PCT
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/00 — A2

(11) International Publication Number: WO 98/24810

(43) International Publication Date: 11 June 1998 (11.06.98)

(21) International Application Number: PCT/EP97/06956

(22) International Filing Date: 3 December 1997 (03.12.97)

(30) Priority Data:
9625283.8 4 December 1996 (04.12.96) GB

(71) Applicant *(for all designated States except US)*: JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B–2340 Beerse (BE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PLATTEEUW, Christ, Jules [BE/BE]; Evergemsesteenweg 17, B–9032 Wondelgem (BE). BUESA ARJOL, Carlos, Manuel [ES/ES]; Travessera de les Corts, 171/702a, E–08028 Barcelona (ES). DERAEYMAEKER, Marc [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). VERHASSELT, Peter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). PUJOL, Nathalie, Jeanne, Raymonde [FR/BE]; 213, avenue du Père Soulas, F–34000 Montpellier (FR). MAERTENS, Luc, Jacques, Simon [BE/BE]; Vier Uitersten 26, B–8200 Brugge (BE). LUYTEN, Walter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). GEERTS, Hugo [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). VANDEKERCKHOVE, Joel, Stefaan [BE/BE]; Rode Boukendreef 27, B–8210 Loppem (BE). GEYSEN, Johan [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B–2340 Beerse (BE). BOGAERT, Thierry, André, Olivier, Eddy [BE/BE]; Wolvendreef 26g, B–8500 Kortrijk (BE).

(74) Agent: BALDOCK, Sharon, Claire; Boult Wade Tennant, 27 Furnival Street, London EC4A 1PQ (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: VERTEBRATE HOMOLOGUES OF UNC–53 PROTEIN OF C. ELEGANS

(57) Abstract

Vertebrate protein homologues of UNC–53 protein of C. elegans and nucleic acid sequences coding for said homologues or functional equivalents thereof are identified. The nucleic acid sequences in an appropriate vector are used to transfect or transform cells, tissues or organisms useful in identifying inhibitors or enhancers of the vertebrate homologue, or further proteins involved in the signal transduction pathway of which said vertebrate homologue is a component. Any of said inhibitors or enhancers identified can be included in a pharmaceutical composition or in the preparation of a medicament for treating conditions such as neurological diseases, acute traumatic injuries and to promote neuronal regeneration and inhibit metastasis or loss of contact inhibition.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

VERTEBRATE HOMOLOGUES OF UNC-53 PROTEIN OF C. ELEGANS

The present invention relates to vertebrate homologues of UNC-53 protein of C. elegans and cDNA sequences coding for said homologues or functional equivalents thereof. The invention also relates to processes for identifying compounds which control cell behaviour, compounds identified and pharmaceutical compositions containing them in addition to processes and assays for identifying disease states in which said gene or protein is dysfunctional.

The control of cell motility, cell shape and directionality of cell outgrowth of axones or other cell outgrowths is an essential feature in the morphogenesis and function of both unicellular and multicellular organisms. The control of these processes is disturbed in a variety of disease states in which, for example, the Receptor Tyrosine kinase (RTK) signal transduction pathways, or the like, or their downstream intra-cellular pathways (which are shared with other extra-cellular receptors, including cell adhesion molecules like N-CAMS and integrins) are overstimulated.

Some cell surface proteins and extra-cellular molecules controlling the directionality and potential of cell migration have been identified, although the processes involved are not generally understood.

It is generally considered that a long-range migration of a cell process (also known as a growth cone extension) is a stepwise event, whereby prior to and after each extension there is the formation of a structure at the leading edge of the cell which senses signals in the environment instructing the cell to either stabilise a cell process extending in a

preferred direction, or to cause a lamellipodium to extend a process in a given direction. Localised stabilisation of the actin cytoskeleton and association with plus end regions of microtubules is a general cell biological process underlying the choice of directional extension. Microtubule binding directing these processes has not previously been identified. The present inventors have surprisingly found that UNC-53 protein of C. elegans and vertebrate homologues thereof is involved in binding of microtubules and particularly of plus end regions of microtubules.

A gene from the free-living nematode Caenorhabditis elegans designated "unc-53" has been previously identified and cloned (Abstract, International C. elegans Meeting, June 1-5 1991, Madison, Wisconsin, 58, Bogaert and Goh). The present inventors previously identified UNC-53 protein as a signal transducer or signal integrator controlling the directionality of cell migration and/or cell shape in C. elegans (WO 96/38555). Increased UNC-53 protein activity was found to be proportional to cell process extensions in the correct direction of cell migration. The unc-53 gene was found to encode a signal transduction molecule that transduces a signal from an RTK such as, for example, via the adaptor protein SEM-5/GRB-2, to the machinery controlling directional growth cone extension or stabilisation, in a highly dosage - dependent fashion.

Genetic and experimental analysis of C. elegans UNC-53 mutants showed that mutations in the unc-53 gene do not affect the general ability of cells to migrate but rather affect the ability of cells to migrate under specific antero-posterior cues. Reduction of UNC-53 activity leads to loss of

direction and reduction of growth cone extension as indicated by the directionality of random extension cycles observed in excretory canal growth cones in UNC-53 mutants.

The function of UNC-53 is highly sensitive to its dosage or activity. Reduction of function leads to proportional reduction of migration to the specific signal while increased expression, using transgenic expression of UNC-53 in muscle cells, leads to increased directional migration. The data lead to the conclusion that UNC-53 functions as an integrator of a directional signal in the organism whereby reception of signals leads to growth cone extension in the correct direction.

Certain alleles of UNC-53 enhance the sex myoblast migration defect of SEM-5 C. elegans mutants in a receptor tyrosine kinase signal transduction pathway (Stern et al 1993 mol. Biol. cell, 4, 1175-1188). While the genetics suggests that UNC-53 and SEM-5 cooperate to regulate sex myoblast migration, genetic experiments do not permit a conclusion that this is the result of a direct molecular interaction. The inventors previously identified a potential sem-5/GRB-2 binding site and showed in two types of biochemical experiments that UNC-53 physically interacts with SEM-5. The present inventors conclude that UNC-53 encodes a signal transduction molecule that transduces extracellular signals for directional migration via the adapter protein SEM-5/GRB-2 to the machinery controlling directional growth cone extension or stabilization.

Several lines of evidence indicate that UNC-53 might act as an adapter linking extracellular signals to the actin cytoskeleton. Firstly, UNC-53 has shown homology to cortical actin binding proteins and that

it is capable of binding F-actin in vitro. In addition expression of UNC-53 in mammalian cells leads to changes in the F-actin cytoskeleton. Very low levels of UNC-53 expression increase the number of filopodia and actin microspikes protruding from the cell surface. Cells expressing UNC-53 also exhibit increased neurite extension and increased cell motility. UNC-53 thus also acts as an activator of migration.

Considering all available data the following possible mechanisms of action of UNC-53 can be formulated.

The choice and activation of directional growth cone extension can be accounted for by local activation of UNC-53 via a SEM-5/GRB2 complex to a receptor (eg receptor tyrosine kinase signal) which reads a localized or directional signal. Changes in growth cone steering are preceded by the formation of a localized actin patch in the area of the growth cone receiving the highest signal (Bentley and O'Connor et al. Curr. Op. NeuroBiol. 1994, vol 4, 43-48). UNC-53 might be directly involved in forming these actin patches through its own actin binding or cross-linking properties. Alternatively activated UNC-53 may (eg via its nucleotide binding domain) transduce a signal to as yet unidentified effectors. For example, activation of the small GTP-binding protein cdc42 or a related protein leads to formation of small actin patches as well as the formation of small filopodia. The unc-53 pathway may be upstream of cdc42 or both signal transducers might share downstream pathways.

The present inventors thus decided to investigate if a similar protein was present in higher organisms such as vertebrates.

The present inventors describe the identification

of a family of genes in vertebrates, and particularly in man and mouse with extensive structural homology to UNC-53. The present inventors have surprisingly found that the nucleotide domains of UNC-53 from C. elegans and UNC-53 from vertebrates similarly activate motility, establishing functional equivalence. Furthermore these domains are shown to be capable of transforming NIH3T3 cells in vitro. The inventors also found changes in RNA transcripts in transformed cell lines compared to normal human tissues suggesting a role for UNC-53 in cell differentiation, morphogenesis and disease. Furthermore, in vitro assays and transgenic models are also described that identify pharmacological modulators of UNC-53 activity and assays to identify proteins interacting with UNC-53.

According to a first aspect of the present invention, there is provided a vertebrate protein homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, which protein homologue comprises an amino acid sequence having a statistically significant homology to the UNC-53 protein of C. elegans as illustrated in figure 2.According to the present invention a derivative should be taken to mean mutational derivatives, fusions, internal deletions, splice variants and muteins.

There is also provided according to a second aspect of the present invention a vertebrate protein homologue of UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence homology blocks A, B, C, D or E as illustrated in Figure 9a, or block F in Figure 12a or a sequence having a statistically significant homology therewith.

Preferably, said vertebrate homologue is a human protein or a mouse protein.

According to a further aspect of the invention there is provided a vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D,E or F which differ from those blocks of Figure 9a and Figure 12a to a significant extent only in conservative amino acid changes. In an even further aspect of the invention there is provided a vertebrate protein having an amino acid sequence encoded by the nucleotide sequence from position 1 to position 6013 as illustrated in Figure 9b. There is also provided a vertebrate protein having an amino acid sequence encoded by the nucleotide sequence illustrated in Figure 11d, or a functional equivalent derivative, or bioprecursor of said homologue.

According to a further aspect of the present invention there is provided a vertebrate protein having an amino acid sequence corresponding to the prosite signatures as illustrated in Figure 28 for each of said homology blocks as defined above. Advantageously the prosite signatures can be used to identify a protein having a statistically significant homology to the UNC-53 protein of C. elegans. (Luethy et al 1994, Protein Science, 3, 139-146).

A further aspect of the invention comprises a vertebrate homologue according to the invention comprising an amino acid sequence as shown in figure 9b or 11d or an amino acid sequence which differs from the amino acid sequences shown in these figures to a significant extent only in one or more conservative amino acid changes.

In a further aspect of the present invention there is also provided a nucleic acid molecule, which

is preferably DNA, and which encodes a vertebrate homologue of UNC-53 protein of C. elegans, or a functional equivalent derivative, fragment or bioprecursor of said homologue according to the invention. Preferably, the cDNA comprises a sequence of nucleotides encoding an amino acid sequence as illustrated in figures 9b or 11d or an amino acid which differs from the sequences shown in these figures to a significant only in one or more conservative amino acid changes. Preferably the DNA is cDNA, which cDNA comprises at least from position 1 to 6013 of the sequence shown in Figure 9b. Alternatively the cDNA may comprise the sequence illustrated in Figure 11d. Also provided by the present invention is a nucleic acid sequence capable of hybridising to the nucleic acid or DNA sequences according to the invention under high strigency conditions, which conditions are well known to those skilled in the art.

The cDNA according to the invention may be included in an expression vector which may itself be used to transform or transfect a host cell, which cell may be bacterial or eukaryotic in origin including such as, for example an animal or plant cell a fungal cell or an insect cell. Thus, advantageously, once the cDNA corresponding to the genome of the vertebrate homologue of UNC-53 of C. elegans is synthesised, using for example, reverse transcriptase or the like, a range of cells, tissues or organisms may be transfected following incorporation of the selected cDNA clone into an appropriate expression vector.The expression vector according to the invention may comprise a promoter of C. elegans or one of human mouse or viral origin and optionally a sequence encoding a reporter molecule, such as, for example,

green fluorescent protein.

The present invention, therefore, also further comprises a transgenic cell, tissue or organism comprising a transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment derivative or bioprecursor of said homologue. The term "transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans" as used herein means a suitable nucleic acid sequence which leads to the expression of a vertebrate homologue of UNC-53 protein of C. elegans having the same function and/or activity. The transgene may include, for example, genomic nucleic acid isolated from the appropriate vertebrate or synthetic nucleic acid including cDNA. The term "transgenic organism, tissue or cell, as used herein means any suitable organism and/or part of an organism, tissue or cell, that contains exogenous nucleic acid either stably integrated in the genome or in an extrachromosomal state.

Preferably the transgenic cell comprises any of, a COS cell, HepG2 cell, MCF-7 or N4 neuroblastoma cell or a NIH3T3 cell or a colorectal or carcinoma cell or a human derived cell such as a fibroblast or the like. The transgenic organism may be an insect, a non-human animal or a plant and preferably C.elegans or a related nematode. Preferably, the transgene comprises the nucleic acid sequence encoding the vertebrate homologue or a functional fragment of said gene according to the invention as described above. The transgene preferably comprises an expression vector according to the invention.

The term "functional fragment" as used herein should be taken to mean a fragment of the gene coding for the vertebrate homologue of the UNC-53 protein of

C. elegans or a functional equivalent or derivative or bioprecursor of said protein. For example, the gene may comprise deletions or mutations but may still encode a functional vertebrate homologue of UNC-53 protein.

Further provided by the present invention is a method of producing a mutant vertebrate non-human organism or cell having a mutation in the wild-type gene coding for the vertebrate homologue of UNC-53 protein, which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes located thereon, which method comprises inducing a mutation in the vertebrate homologue of UNC-53 protein in said organism or cell. These mutant organisms or cells may be used in a screen to identify the effects of compounds on these cell functions.

The vertebrate homologue of UNC-53 protein of C. elegans or the cDNA or genomic DNA encoding it or a functional equivalent, derivative, fragment or bioprecursor of said homologue, may advantageously be used as a medicament, or in the preparation of a medicament to promote neuronal regeneration, revascularisation or wound healing or the treatment of chronic neurodegenerative disorders or acute traumatic injuries or fibrotic disease or physiological events requiring the polarity of cells or epithelia. The present inventors have also found that the vertebrate homologue of UNC-53 protein plays a role in a transformed state of cells. Accordingly, the vertebrate homologue, dominant positive or negative mutants thereof, or inhibitors thereof may advantageously be used to induce or alleviate contact inhibition in a cell or in preventing cancer

development. Typically, the above medical conditions may be treated in mammals and more preferably humans by either a homologue of UNC-53 protein or alternatively by a nucleic acid coding for such a protein. Alternatively an antisense oligonucleotide to said UNC-53 homologue may be used to prevent it's expression. Examples of other nucleic acid sequences which may be used include 3' untranslated regions of mRNA which could be used to prevent transcription of the genomic sequence encoding for the vertebrate homologue of UNC-53 protein.

The vertebrate homologue of UNC-53 protein or a functional equivalent, fragment or bioprecursor of said protein may be incorporated into a pharmaceutically acceptable composition together with a suitable carrier, diluent or excipient therefor. The pharmaceutical composition may advantageously comprise, additionally or alternatively, the nucleic acid sequence according to the invention as defined above.

The present invention also provides for a method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, transformation, cell shape or motility or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes thereon which method comprises contacting said compound with a transgenic cell according to the invention and screening for a phenotypic change in said cell. Preferably the method can determine whether the compound comprises an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which pathway said vertebrate homologue of UNC-53 protein, or a functional equivalent, derivative, fragment or bioprecursor of

said vertebrate homologue is a component or whether said compound is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell. The present invention also provides a method to determine that the protein in said signal transduction pathway is a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

Preferably, the phenotypic change to be screened comprises a change in cell shape or a change in cell motility. Where a transgenic cell is used in accordance with one embodiment of the method of the invention, an N4 neuroblastoma cell may be used and in such an embodiment the phenotypic change to be screened may be the length of neurite growth or changes in filipodia outgrowth or alternatively changes in ruffling behaviour or cell adhesion or any change in microtubule cytoskeleton or any change in localisation of proteins on plus end regions of microtubules or any change in cell death such as in apoptosis. In an alternative embodiment of the method of the invention, the transgenic cell may comprise an MCF-7 breast cancer cell. Typically in such an embodiment the phenotypic change to be screened comprises the extent of phagokinesis or filipodia formation. In an alternative embodiment of this aspect of the invention, the transgenic cell may comprise an NIH3T3 cell. Typically in such an embodiment the phenotypic change to be screened comprises loss of contact inhibition of foci formation. The method according to the invention, may also utilise a mutant cell or mutant organism according to the invention as described above, where the mutant cell is capable of growing in tissue

culture or in vivo and either of which cell or organism has a mutation in the wild-type unc-53 gene.

In accordance with the present invention, a "phenotypic change", may be any phenotype resulting from changes at any suitable point in the life cycle of the cell, tissue or organism defined above, which change can be attributed to the expression of the transgene such as for example, growth, viability, morphology, behaviour, movement, cell migration or cell process or growth cone extension of cells and includes changes in body shape, locomotion, chemotaxis, contact inhibition, mating behaviour or the like. The phenotypic change may preferably be monitored directly by visual inspection of the cell as a whole or particularly by monitoring the F-actin cytoskeleton microtubule network and plus end stability of microtubules or proteins thereon or alternatively by for example measuring indicators of viability including endogenous or transgenically introduced histochemical markers or other reporter genes, such as for example ß-galactosidase or green fluorescent protein.

A compound which is identifiable by the method according to the invention as described above, as an enhancer of the processes identified above such as the regulation of cell shape or motility or the direction of cell migration may be used as a medicament, or alternatively in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Examples of promoting neuronal regeneration include, for example, peripheral nerve regeneration after trauma and spinal cord trauma.

Where a compound is identified in accordance with

the method described above as being an inhibitor of the regulation of cell shape etc., the compound may be used as a medicament, or in the preparation of a medicament, for substantially alleviating spread of disease inducing cells, such as in spread of cancer, or the like in metastasis or in alleviating loss of contact inhibition. Advantageously, any of the compounds which may have been identified as an inhibitor or an enhancer in accordance with the method as described above, may also be included in a pharmaceutical composition comprising the respective compound and a pharmaceutically acceptable carrier, diluent or excipient therefor.

The particular mechanism of action of a compound identified as either an inhibitor or an enhancer of the cell motility shape, growth or direction of cell migration or microtubule association or to the plus end region thereof is not limiting. Preferably the compound acts as an inhibitor or enhancer of a signal transduction pathway. The compound may also act on a parallel pathway or directly on the vertebrate homologue of UNC-53 protein of C. elegans. For example, the method of action of the compound may include direct interaction with the vertebrate homologue of UNC-53 protein, interaction with processes for regulating phosphorylation or dephosphorylation of the vertebrate homologue of UNC-53 or with processes regulating activity of an unc-53 gene or with processes for post-transcriptional or post-translational modification or the like.

Preferably the compound is identified by the method according to the invention as an inhibitor or an enhancer, by utilising differences of phenotype of the cell, tissue or organism, which are visible to the eye. Alternatively indicators of viability including

endogenous or transgenically introduced histochemical markers or a reporter gene may be used.

According to a further aspect of the invention there is also provided a transgenic cell or tissue culture which has been constructed to comprise a promoter sequence of a gene coding for a vertebrate homologue of UNC-53 of C. elegans or a functional equivalent, derivative fragment, or bioprecursor of said homologue operably linked to a nucleic acid sequence encoding a reporter molecule. Preferably, the reporter sequence encoding the reporter molecule which comprises a detectable protein, for example one which may be monitored by eye inspection such as antibiotic resistance, ß-galactosidase or a molecule detectable by spectrophotometric, spectrofluorometric, luminescent or radioactive assays.

The present invention also provides a method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein in C. elegans, or a functional equivalent, derivative fragment or bioprecursor of said homologue, which method comprises the steps of:

(a) contacting said compound with a transgenic cell according to the invention as described above,

(b) monitoring the level of said reporter molecule and comparing results obtained from this monitoring step with a control comprising a transgenic cell having the promoter sequence of a gene coding for a vertebrate homologue of UNC-53 protein, or a functional fragment of said homologue and the reporter molecule, in the absence of the compound.

In one embodiment of the method according to this

aspect of the invention the reporter molecule may comprise messenger RNA.

A compound identified as an enhancer of transcription of the gene coding for the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor of said homologue may also be used as a medicament, or in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Furthermore, such compounds may be included in a pharmaceutical composition including a pharmaceutically acceptable carrier, diluent or excipient therefor. Any compounds identified as inhibitors of transcription may, advantageously, be used in alleviating the spread of disease inducing cells such as cancers or metastasis or loss of contact inhibition.

The present invention also provides a kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell growth, transformation, cell motility or shape or the direction of cell migration which kit comprises at least one transgenic or mutant cell or transgenic or mutant non-human organism according to the invention as described above and a plurality of wild-type cells or one organism of the same type, or a cell line or tissue culture and means for contacting said compound with said cell or organism.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or fragment

thereof, which kit comprises at least one transgenic cell or cells according to the invention and means for contacting said compounds with said cells.

For the purposes of the present invention, the term "gene coding for a vertebrate homologue of UNC-53 or a functional fragment of said homologue" includes the nucleic acid sequence shown in Figures 9b or 11d or a fragment thereof, including the differentially spliced isoforms and transcriptional starts of the nucleic acid sequence and which sequence encodes a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor of the protein.

The present invention also provides methods of identifying genes of vertebrates or fragments of said genes, which encode proteins which are active in the signal transduction pathway of which the vertebrate homologue of UNC-53 is a component. A preferred method comprises hybridizing to an appropriate cDNA library a nucleotide sequence, as defined herein, or a fragment thereof under appropriate conditions of stringency in order to identify genes having statistically significant homology with the cDNA clones of any one of the cDNA sequences according to the invention described above.

Furthermore, there is also provided by the present invention a method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said vertebrate homologue is a component. According to this aspect of the invention, the method comprises;

(a) contacting an extract of said cell with an antibody to the vertebrate homologue of UNC-53

protein or a functional equivalent, fragment or bioprecursor of said protein,

(b) identifying the antibody/vertebrate homologue of UNC-53 complex, and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the antibody.

The vertebrate homologue of UNC-53 protein, therefore may bind regions of other proteins involved in the signal transduction pathway. It is also possible to sequentially identify a whole range of proteins involved in the signal transduction pathway.

Antibodies to the vertebrate homologue of UNC-53 protein may be produced according to known techniques as would be known to those skilled in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with a protein or epitope of a protein according to the invention and recovering immune serum.

This aspect of the invention further comprises a method of identifying a further protein or proteins which are active in the signal transduction pathway of a cell of which UNC-53 is a component which method comprises:

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein in the method as described above,

(b) contacting a cell extract with the antibody,

(c) identifying the antibody/protein complex,

(d) analysing the complex to identify any further protein bound to the first protein other than the antibody, and

(e) optionally repeating steps (a) to (d) to identify further proteins in the pathway.

According to this aspect of the present

invention, the antibody starts the process by binding to the vertebrate homologue of UNC-53 protein or a functional equivalent thereof in the signal transduction pathway. Any other proteins found complexed to the bound antibody or UNC-53 protein can then be used to identify further interacting proteins involved in the pathway.

It may also be possible to identify proteins involved in the signal transduction pathway of a cell of which the vertebrate homologue of UNC-53 or a functional equivalent derivative or bioprecursor thereof is a component by using a vertebrate homologue of UNC-53 protein of C. elegans. According to this aspect of the invention the method comprises:

(a) contacting an extract of the cell with the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue,
(b) identifying the vertebrate homologue of UNC-53 protein/protein complex formed and
(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein

This method can also advantageously be used to identify further proteins in a signal transduction pathway of a cell by contacting an extract of the cell used as described above, with any protein identified from step (c) above not being a vertebrate homologue of UNC-53 protein and repeating steps (b) and (c).

Other methods which may be used for identifying proteins in a signal transduction pathway of a cell may comprise for example a western blot overlay method which method is well known to those skilled in the art. Cell extracts are run on gels to separate out

protein and subsequently blotted onto a nylon membrane. These membranes may then be incubated, for example in a medium containing a vertebrate homologue of UNC-53 having a label attached thereto such as a biotin or radiolabel and any protein conjugates visualised with for example a streptavidin or alkaline phosphatase conjugated antibody.

The present invention also advantageously provides a process for the preparation of binding antibodies which recognise proteins or fragments thereof involved in the rate and direction of cell migration or the control of cell growth or shape, for the above methods.

The monoclonal antibody for binding to the appropriate vertebrate homologue of UNC-53 (or its functional equivalent) may be prepared by known techniques as described by Kohler R. and Milstein C., (1975) Nature 256, 495 to 497.

Another method which may be used to identify proteins involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent or derivative or bioprescursor is a component involves investigating protein-protein interactions using the two-hybrid vector method. This method is well known to those skilled in the art and which was first developed in yeast by Chien *et al* (1991). This technique is based on functional reconstitution in vivo of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA

sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as a library or the like,encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example a sequence coding for the vertebrate homologue of UNC-53. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein, preferably from the signal transduction pathway of the vertebrate in question. Any interaction between the vertebrate homologue of UNC-53 protein and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter

molecule such as ß-galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes. This method enables any interactions between proteins involved in the signal transduction pathway or a parallel or redundant pathway to be investigated.

Any proteins identified in the signal transduction pathway of the cell, which may be for example a mammalian cell, may also be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

The present invention also provides a process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, or derivative of the protein, which process comprises culturing the cells transformed or transfected with a cDNA expression vector having any of the cDNA sequences according to the invention as described above, and recovering the expressed vertebrate homologue of UNC-53 protein. The cell may advantageously be a bacterial, animal, insect or plant cell.

A particularly preferred process for producing a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or fragment of said homologue comprises using insect cells. Accordingly, the invention provides a process for producing a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of the UNC-53 protein, which process comprises culturing an insect cell transfected with a recombinant Baculovirus vector, said vector comprising a nucleotide vector encoding the vertebrate homologue of UNC-53 protein or a functional equivalent, fragment

or bioprecursor thereof downstream of the Baculovirus polyhedrin promoter and recovering the expressed protein. Advantageously, this method produces large amounts of protein for recovery. The insect cell may be from for example Spodoptera frugiperda or Drosophila Melanogester.

In accordance with the present invention, a defined nucleic acid sequence includes not only the identical nucleic acid but also any minor base variations from the natural nucleic acid sequence including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid), due to the degenerate code in conservative amino acid substitution. The term "nucleic acid sequence" also includes the complimentary sequence to any single stranded sequence given which includes the definition above regarding base variations.

Furthermore, a defined protein, polypeptide or amino acid sequence according to the invention, includes not only the identical amino acid sequence but also minor amino acid variations from the natural amino acid sequence including conservative amino acid replacements (a replacement by an amino acid that is related in its side chains). Also included are amino acid sequences which vary from the natural amino acid but result in a polypeptide which is immunologically identical or similar to the polypeptide encoded by the naturally occurring sequence. Such polypeptides may be encoded by a corresponding nucleic acid sequence.

A further aspect of the invention provides a nucleic acid sequence of at least 15 nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides.

These sequences may, advantageously be used as

probes or primers to initiate replication or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting for the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with a sample under hybridising conditions and detecting for the presence of any duplex formation between the probe and any nucleic acid in the sample. Nucleic acid sequences according to the invention may also be produced using recombinant or synthetic means such as described in Sambrook et al (Molecular Cloning: A Laboratory Manual,1989).Advantageously, human allelic variants or polymorphisms of the DNA according to the invention may be identified by, for example, probing DNA libraries from a range of individuals for example from different populations. Furthermore, nucleic acids and probes according to the invention may be used to sequence genomic DNA from patients using techniques well known in the art, such as the Sanger Dideoxy chain termination method, which may advantageously ascertain any predisposition of a patient to certain proliferative disorders.

A method of detecting whether a compound is an inhibitor or an enhancer of expression of a vertebrate homologue of UNC-53 of C. elegans, or a functional equivalent, derivative or fragment of said vertebrate homologue is also provided which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

Preferably the cell is a transgenic cell as described above. Alternatively the cell may have

undergone loss of contact inhibition.

The present method also provides for determining whether said compound is an inhibitor of expression of said vertebrate homologue. In one embodiment the compound to be tested comprises a nucleic acid.

Preferably said nucleic acid sequence comprises an antisense DNA sequence or a mRNA sequence.

Preferably said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

Alternatively,the compound to be tested may be a protein. Preferably, said protein comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue and preferably comprises a protein identified by the methods as described herein.

The present invention also provides a pharmaceutical composition comprising a compound, for example an antisense nucleic acid identified according to the above described method together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

A nucleic acid sequence or protein identified according to this aspect of the invention may be used as a mediciament, or in the preparation of a medicament, for treating loss of contact inhibition or cancer which is mediated by a vertebrate homologue of UNC-53 protein or a functional equivalent, fragment, derivative or bioprecursor of said homologue.

Further provided by the invention is a nucleic acid as defined above for use in preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said homologue.

According to a further aspect of the invention there is provided a plasmid pCB201 deposited under LMBP Accession No. LMBP 3594 and a MCF-7 and a NIH/3T3 cell line transfected with plasmid pCB201 deposited under LMBP Accession Nos. LMBP 1601 CB and LMBP 1603 CB respectively. Further provided by the invention is phage lambda 3b coding for Hu-UNC-53/1 and deposited under Accession No. LMBP 1604CB (or 3595). Also provided are plasmids pLM1 deposited under Accession No. LMBP 3762, pLM4 (LMBP 3763), pEGFP72 (LMBP 3764) and pCB501 (LMBP 3765). Further provided is a Bac clone comprising a fragment of hu-unc-53/2 gene (LMBP 3773) and a worm strain comprising a chimeric C.elegans human unc53 gene deposited under LMBP Accession No. LMBP-1663CB.

Further provided by the invention is an assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, or a functional equivalent, derivative or bioprecursor thereof, which antibody is fused to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

Preferably the reporter molecule is an antibody conjugated to for example a flurophore such as fluorescein or alternatively to an enzyme such as strepavidin.

There is also provided a method for detecting for expression of a gene coding for a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor thereof, which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue or a

functional equivalent, fragment or bioprecursor thereof with a cell extract, which probe is linked to a reporter and analysing for the presence of said reporter.

Preferably the probe is a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or bioprecursor therefor.

Preferably the complimentary sequence is a 3' or 5' untranslated region of said mRNA. Preferably said reporter may be a dig label, a fluorophore, a hapten or a radiolabel.

Alternatively said probe comprises an antibody specific for said vertebrate homologue of said UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor therefor.

Preferably the reporter is an antibody conjugated to for example a fluorophore such as fluorscein or alternatively an enzyme such as streptavidin.

As described above UNC-53 protein of C.elegans has been found to localise to microtubule and particularly to microtubule (+) ends. Therefore, there is provided by a further aspect of the present invention a method of determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to any of claims to 1 to 9 to microtubules or plus end regions thereof, which method comprises (a) contacting said compound with a transgenic cell, tissue or organism expressing UNC-53 protein or said vertebrate homologue and which protein is operably linked to a reporter molecule (b) screening for the localisation of said reporter molecule as compared to a cell according to step (a) which has not been contacted

with said compound.

A compound identifiable by the above method also forms part of the present invention. Such a compound identified as an inhibitor of localisation or association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof may be used in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition. Further a compound identified as an enhancer of association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof may be used in for example promoting neuronal regeneration, revascularisation or wound healing, or for treating chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease. These compounds may then be included in a pharmaceutical composition, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to the invention with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing UNC-53 and a reporter molecule or a host or transgenic cell according to the invention and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells. Compounds identified as inhibitors or enhancers or microtubule association described above may advantageously be included in a composition and linked to unc-53 protein of C.elegans or a vertebrate homologue thereof according to the invention to target the compounds to the microtubules or the plus end regions thereof. Such a composition

may also comprise, for example, a suitable transfecting or transformation agent.

According to a further aspect of the invention there is provided a method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable of expressing UNC-53 or a vertebrate homologue thereof according to the invention, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting said protein to said microtubule or a plus end region thereof. An even further aspect of the invention comprises a method of identifying a molecule which covalently modifies UNC-53 or a vertebrate homologue thereof according to the invention, which method comprises a) contacting either an extract from a cell or cells expressing UNC-53 or said vertebrate homologue or a mixture of enzymes comprising canditate UNC-53 modifying enzymes in the presence of an indicator of covalent modification of a protein, b) identifying any covalently modified UNC-53 protein from step a)and c) identifying said molecule involved in said modification step. Such an indicator may be ^{32}p.

Further provided by the invention is a method of identifying a compound which alleviates or enhances the toxicity of UNC-53 or a vertebrate homologue thereof according to the invention, or which alleviates or enhances apoptosis. The method of the former comprises contacting said compound with a transgenic cell, tissue or organism according to the invention and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof.In the case of apoptosis the

method comprises monitoring the effect of the compound on cell death.

The invention may be more clearly understood from the following examples which are only exemplary, with reference to the accompanying drawings wherein

Figure 1 illustrates the sequence of plasmid pTB72 which codes for the full length UNC-53 protein in C. elegans,deposited under LMBP Accession No. 3486.

Figure 2 illustrates the full-length UNC-53 protein from C. elegans.

Figure 3 is a Tblastn search of the EST division of Genbank with the ORF of the longest known Ce-UNC-53 cDNA. tb3-M5, reveals two EST's with homology to a predicted coiled-coil region in Ce-UNC-53.

Figure 4 illustrates a search of the Genbank databases with part of the nucleotide binding domain of Ce-UNC-53. It does not identify statistically significant proteins except for the C. elegans cosmid containing Ce-unc-53.

Figure 5 illustrates a three frame translation of EST gb:R41071.

Regions of homology with Ce-Unc-53 in two different frames are underlined. The spacing between the blocks of homology is of similar size to that in Ce-UNC-53. Subsequent re-cloning and re-sequencing of this region in man identified multiple sequencing errors gb:R41071, and identified an ORF which is more homologous to and co-linear with Ce-UNC-53 (see alignment in fig. 12).

Figure 6 is a BLASTN search of the EST division of Genbank with Hu-unc-53/1 cDNA cosmid 3b.

Figure 7 is a TBLASTN search of the Genbank sequence database with the 961 amino acid ORF of cDNA 3b of hu-UNC-53/1 : hu-UNC-53/1 forms a unique pair with Ce-UNC-53 (cosmid F45E10) compared to the rest of

the database.

Figure 8 is a diagram illustrating the length and overlap and tissue source of the different cDNA clones of the 3' end of Hu-UNC-53/1 isolated in this work.

Figure 8a. is a diagram illustrating the further sequence of the Hu-UNC-53/1 and overlap of constructs to obtain the further sequence.

Figure 8b is a diagram illustrating the 3' end of Hu-UNC53/1 and the EST clones present in the database.

Figure 9a is an annotated sequence listing of clone 3b of hu-UNC-53/1 including the EcoR1 polylinker GAATTC. The predicted Open Reading Frame of Hu-UNC-53/1 is listed below the sequence. Blocks A B C D and E which are similiar to Ce-UNC-53/1, a region which is different between Hu-UNC-53/1 and Hu-UNC-53/2 and the 3' untranslated leader sequences are marked with arrows and labelled.

Figure 9b is an annotated sequence listing of Hu-UNC-53/1 available at this moment. The predicted Open Reading Frames of Hu-UNC-53/1, pLM1, pLM3, pLM4, pCB251, pLM5 and pCB201, the homology blocks A,B,C,D and E, the position of a region which is different between Hu-UNC-53/1 and Hu-UNC-53/2, the position of phh14-3, pCB212, pCB210-14, phh3b, phh15, the position of the reverse primers HU53rv1, HU53rv2, HU53rv3 and HU53rv4, the position of peptides B72628 (=28/1), B72627, B72626 and B72625 are listed below the sequence.

Figure 10 is an annotated sequence listing of the insert of clone gbAA049124 (EST479167) of mu-UNC-53/1. The open reading frame and 3' untranslated sequence is marked with an arrow.

Figure 11a is an annotated sequence listing of the insert of clone gbH09036 (EST46037) of Hu-UNC-

53/2.

Figure 11b is a novel DNA sequence of HU-UNC-53/2 extended by RT-PCR. This DNA sequence is not present in EST-46037 and extends the ORF beyond position 1109 of Figure 11a to an ORF from position 18 to 1793.

Figure 11c summarises how the 3' and 5' extensions of hu-unc-53/2 were made.

Figure 11d compiles the sequence of hu-unc-53/2. The boxed sequences are the primer sequences used for the respective extension steps described in the experimental methods section.

Figure 11e illustrates the sequences of the extensions summarised in figure 11c.

Figure 11f illustrates the sequence information illustrating four alternative Start sites observed for hu-unc-53/2.

Figure 12.is an illustration of a Tblastn search of the EST division of Genbank with 680aa starting at the C-teminus of the alpha actinin domain of hu-unc-53/2.

Figure 12a.is an illustration of an amino acid alignment of the available sequence of C.elegans unc-53 and hu-unc-53/1 and hu-unc-53/2.

12b. is an illustration of similarity plots for Ce-unc-53 and hu-unc-53/1 (top) and for hu-unc-53/1 and hu-unc-53/2.

Figure 13 is an annotated sequence listing of expression vector pCB201 containing homology block E from Hu-UNC-53/1 cloned in a pcDNA3.1-HIS expression vector. The HIS and T7-tags, PCR primer used to modify hu-UNC-53/1 and ORF are marked.

Figure 14 is a diagram showing the alignment of the homologous regions of hu-UNC-53/1 and mu-UNC-53/1.

Figure 15 is an annotated sequence listing of expression vector pCDU3 containing part of Ce-UNC-53/1

cloned in expression vector pcDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

Figure 16 is an annotated sequence listing of expression vector pCDU4 containing part of Ce-UNC-53/1 cloned in expression vector pcDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

Figure 17 is an annotated sequence listing of expression vector pCDU2 containing part of Ce-UNC-53/1 cloned in expression vector pcDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

Figure 18 illustrates MCF-7 cells transfected with pCB201 (upper) compared to mock transfected MCF-7 cells (phase contrast image). The control cells are spread out on the tissue culture plastic and exhibiting few filopodia outgrowths. The transfected cells appear smaller because they are slightly rounded up and have multiple filopodia outgrowths per cell (arrowheads).

Figure 19 is a phase contrast image of MCF-7 cells, transfected with pcDNA3.1 (19a), pCDU4 (19b), pCDU3 (19c), pCDU2 (19d) and pTB72 (19e).

Figure 20 is an F-actin pattern (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3.LacZ (top panel) and with pCB201 (middle and lower panel).

Figure 21 is an F-actin pattern Phalloidin (visualised with TRITC-Phalloidin) of MCF-7 cells transfected with pCDNA3.1 (21a), pCDU4 (21b), pCDU3(21c), pCDU2 (21d) and pTB72 (21e).

Figure 22 is a phase contrast image of N4 neuroblastoma cells transfected with pcDNA3.1 (22a), pCDU4 (22b), pCDU3 (22c), pCDU2 (22d) and pTB72 (22e).

Figure 23 is an F-actin pattern Phalloidin (visualised with TRITC-Phalloidin) of N4 neuroblastoma cells transfected with pcDNA3.1 (23a), pCDU4 (23b), pCDU3 (23c), pCDU2 (23d) and pTB72 (23e).

Figure 24 illustrates phase contrast images of small (top), medium (middle) and large foci (bottom) induced in a monolayer of NIH3T3 cells by transfection with pCB201.

Figure 25(c) illustrates human metaphase chromosomes probed with a probe 1p34 and figures 25a and 25b indicating the chromosomal location of hu-UNC-53/1 in 1q31. Essentially the same techniques were used to assign the gene hu-unc-53/2 to chromosome locus 11p15 (25d and e) as illustrated in micrograph 25f.

The ideograms 25a and 25d are from the International System for Human Cytogenic Nomenclature 1985. The ideograms 25b and 25e in which the relative band positions and arm ratios were derived from actual chromosome measurements is from Cytogenet Cell Genet 65:206-219 (1994).

Figure 26 is an expression pattern of HU-Unc53/1 and HU-Unc532 in normal human tissues and cancer cell lines.

Figure 27 is a sequence map of Plasmid pNP3.

Figure 28 is an examplary list of prosite signatures which can be used to define and identify vertebrate homologues of UNC-53.

Figure 29 is a annotated sequence map of plasmid pEGFPsac. The GFP-C.elegans unc53sac fusion protein, and the C.elegans unc53 sac fragment are indicated.

Figure 30 is a sequence map of plasmid pEGFP72. The GFP-C.elegans unc53 fusion protein and the C.elegans unc53 fragment are indicated.

Figure 31 is an annotated sequence map of plasmid

pEGFPsma. The GFP-C.elegans unc53sma fusion protein, and the C.e.unc53 sma fragment are indicated.

Figure 32 is an annotated sequence map of plasmid pEGFPecl. The GFP-C.elegans unc53ecl fusion protein, and the C.elegans unc53 ecl fragment are indicated.

Figure 33 is an annotated sequence map of plasmid pEGFPxba. The GFP-C.elegans unc53xba fusion protein, and the C.elegans unc53 xba fragment are indicated.

Figure 34 is an annotated sequence map of plasmid pLM4. Open reading frames of the hul-unc53/1 and GFP are indicated.

Figure 35 is a sequence map of plasmid pNP8.

Figure 36 is an illustration of microtubule association of C.elegans Unc53, shown in HepG2 cells, transiently transfected with pTB72, expressing C.elegans Unc53. panel A:microtubule staining of HepG2 cells, using YL1/2 panel B:C.elegans Unc53 staining, using rab4.

Figure 37 is an illustration of microtubule plus-end association in human cell lines transiently transfected with pTB72, expressing C.e.Unc53. C.elegans Unc53 was stained with mab-16-48. Panel C: COS cells showing microtubule association panel B: MCF7 cells showing microtubule plus-end association panel A: HepG2 cells showing microtubule plus-end association.

Figure 38 is an illustration of microtubule association in N4 cells transiently transfected with pEGFP72, expressing the GFP-C.elegans Unc53 fusion protein. GFP fluorescence was observed in living cells. Panel A: microtubule association of the GFP-C.elegans unc53 fusion protein panel B: microtubule plus-end association of the GFP-C.elegans unc53 fusion protein.

Figure 39 is an illustration of microtubule

association in N4 cells transiently transfected with pEGFP72, expressing the GFP-C.elegans Unc-53 fusion protein. Microtubules were stained with YL1/2 after paraformaldehyde fixation. Panel A: Microtubule association of the GFP-C.elegans unc53 fusion protein. Panel B: tubuline staining. Panel C: panel A plus panel B: co-localisation of the GFP-C.elegans unc-53 fusion protein and Tubuline can be seen as yellow.

Figure 40 is an illustration of microtubule association in N4 cells, transiently transfected with pEGFPsma, expressing the GFP-C.elegans unc53sma fusion protein. Panel A: Microtubule association of the GFP-C.elegans unc53sma fusion product. Panel B: Centriole association of GFP-C.elegans unc53sma fusion product when expressed at low levels.

Figure 41 is an illustration of microtubule association in N4 cells, transiently transfected with pEGFPecl, expressing the GFP-C.elegans unc53ecl fusion protein. Panel A: Microtubule association of the GFP-C.elegans unc53ecl fusion product. Panel B: Centriole association of GFP-C.elegans unc53ecl fusion product when expressed at low levels.

Figure 42(a)/Figure 42(b) are illustrations of fluorescence of GFP in N4 cells transiently transfected with pEFPxba and pEFGPsac respectively.

Figure 43 is an illustration of microtubule association of in N4 cells transiently transfected with pLM4 expressing GFP-Hu-UNC53/1 fusion protein. Panel A: microtubule association of GFP-HU-UNC53/1 fusion protein. Panel B: microtubule plus-end association of GFP-Hu-UNC53/1 fusion protein. Panel C: microtubule association of GFP-Hu-UNC53/1 in dividing cells (end of division).

Figure 44 is an illustration of the sequence of Plasmid pNP9.

Figure 45 is an illustration of immuno fluorescence in melanoma G361 cells stained with sera 28.1. Panel A: Microtubule plus-end association of Hu-UNC53/1. Panel B: microtubule plus-end association of hu1-Unc53 in growth cone extensions.

Figure 46 is an illustration of GFP fluorescence and immunofluorescence in N4 cells transiently transfected with pLM4, and stained with sera 28.1. Panel A: Fluorescence of GFP-Hu-UNC53/1 fusion protein. Panel B: Immunofluorescence of serum 28.1.

Figure 47 is an overview of the microtubule (+) end, the microtubule and f-actin cytoskeleton binding properties of different constructs

Figure 50 is an illustration of rescue of lateral ALN neurons in mutant unc-53.

Dorsal view of the ALN neurones axones visualise in GFP fluorescence with the transgene pA/GFP in the posterior of an adult, (c) cellular body.

a) wild type, anterior axon (aa) migrates in a straight line along the body until reaching the head, on the dorsal sublateral cord, posterior axon (ap) migrates into the tail;

b) *unc-53(n152)*, anterior axons are the shorter, stop ahead of the vulva region and form numerous collateral branches towards the dorsal cord;

c) *unc-53(n152),pA/unc-53* anterior axons no longer form branches and migrate in a straight line into the head, as in the wild type at a).

scale bar 10 μm.

Figure 51a : is an illustration of chimeric fusion between C. elegans and human 1 homologue of the unc-53 gene. The region of the putative nucleotide binding domain (NTP) is replaced in the C. elegans cDNA by the same region of the human homologue 1 of unc-53 (H1). The cDNA is under the promotor region A

(pA) of unc-53, which raise expression in the ALN lateral neurons.

Figure 51b : is an illustration of the chimeric minigene nematode/human pA/unc-53-H1 partially rescue the defect in the longitudinal migration of the lateral neurons ALN and PLN. The four strains compared are : wt; unc-53(n152); unc53 (n152),pA/unc-53; unc-53(n152),pA/unc-53-H1. The observed phenotypes are put in three classes :

½sauvage¶, the axon is straight, unbranched, and migrates until the head; ½vulve¶, the axon is straight, unbranched, and stops in the vulva region; ½mutant¶, the axon is short, never joints the vulva region and made a lot of collateral branches. Numbers are in percentage. The number of observed axons are noted in the last column. The chimeric fusion between the C. elegans gene and human homolog (unc-53-H1) partially rescues the mutant phenotype. The chimeric gene was maded by replacing the putative nucleotide binding region (NTP)of the nematode cDNA by the same region of the human homolog 1 (H1).

Figure 52 is an illustration of the sequence for plasmid pLM5.

Figure 53 is an illustration of the sequence for plasmid pLM6.

Figure 54 is an illustration of the sequence for plasmid pLM1.

Figure 55 is a sequence map of plasmid pCB251.

Figure 56 is a sequence map of plasmid pNP10.

Figure 57 is a sequence map of plasmid pCB501.

Figure 58 is a sequence map of plasmid pTB115.

Figure 59 is a sequence map of plasmid pPD95.75.

Figure 60 is a sequence map of clone X16.

Figure 61 is a sequence map of plasmid pLM3

DEPOSITED MATERIALS

Deposit	Date	Acc. Nr
pCB201 plasmid DNA in E. coli	3 December 1996	LMBP 3594
Lambda clone 3B encoding hu-unc-53/1	3 December 1996	LMBP 3595
MCF-7 clone z4 (mock)	3 December 1996	LMBP 1600CB
MCF-7 clone (pCB201)	3 December 1996	LMBP 1601CB
NIH-3T3 mock	3 December 1996	LMBP 1602CB
NIH-3T3 pCB201	3 December 1996	LMBP 1603CB
pLM1	13 November 1997	LMBP 3762
pLM4	13 November 1997	LMBP 3763
pEGFP72	13 November 1997	LMBP 3764
pCB501	13 November 1997	LMBP 3765
BAC clone comprising fragment of hu-unc53/2 gene	15 November 1997	LMBP 3773
Worm strain with chimeric C.elegans/human unc-53 gene	15 November 1997	LMBP-1663CB

The above plasmids and cell lines were deposited at the Belgian Coordinated Collections of Microorganisms (BCCM) at laboratorium voor moleculaire biologie - plasmidencollective (LMBP) B9000, GENT, Belgium, in accordance with the provisions of the Budapest Treaty of 28 April 1977.

The present invention will now be described with reference to the following examples which are not limiting.

Identification of a human homologue of the UNC-53 protein of C.elegans.

Extensive searches with the ce-UNC-53 sequence (Figures 1 and 2) against the public domain databases

(EST, Genbank, EMBL, Swissprot and PIR) revealed no statistically significant homologies (a smallest sum probability (ssp) of 10 e - 8 is generally accepted to be significant at amino acid level). Two ESTs gbH09036 (ssp = 1.1 e - 5) a Homo sapiens cDNA clone and gbAA049124 (ssp=8.6-5) a mouse cDNA clone showed homology to a "coiled coil" region a common motif in the contributing to protein secondary structure. (figure 3)

All other candidate scores were are at background level (ssp >0.21). Careful examination of weak candidate ESTs identified EST gb:R41071 from Homo sapiens, which had obtained a low score of 53 and a non-significant probability score of 0.33 (Fig. 4). The inventors surprisingly discovered potentially significant homology with the Ce-UNC-53 nucleotide binding domain, provided multiple frameshifts and sequence errors were hypothesized.

The inventors amplified, cloned and sequenced part of gb:R41071 from human heart and human lung cDNA and from human genomic DNA and discovered that clone gb:R41071 had up to ten 10 different mistakes in the region checked. 5 extra nucleotides were scattered along its sequence and two nucleotide substitutions were identified, and gb:R41071 lacked three nucleotides present in our clone (Fig. 5). The novel sequence obtained was two nucleotides shorter and showed the two UNC-53-homologous regions in frame. The genomic fragment obtained is larger (700 bp total length) than the corresponding cDNA clones indicating the presence of an interverting sequence of around 500 bp in nucleotide 162 of this fragment. The amplified cDNA fragment which was cloned to vector PCRII (Intvitrogen) and named pCR231 and was used as a probe to screen cDNA libraries.

The conceptual translation of the clones we obtained by PCR were screened using blast and tblastn against all known protein and DNA sequences in the database. The only clone which came up with statistically significant similarity was Ce-UNC-53 (Fig.6). This human clone and Ce-UNC-53 thus form a unique homologous pair compared to the rest of the known sequences, indicating the statistical relevance and novelty of our discovery. We designate this human gene as hu-UNC-53/1. Human heart and a human colorectal adenocarcinoma cDNA libraries were probed with pCR231 probe to identify longer cDNA clones. The clones overlap giving a linear sequence of 3706 bp (Fig 8 and 26). This sequence shows an 959 amino acid open reading frame from the beginning of the clone. The absence of a 5' untranslated region suggests that the mRNA will extend 5'.

Sequence alignment searches of the public domain databases with the DNA sequence of hu-UNC-53/1 and its' conceptual translation identified a series of ESTs most of which correspond to the 5' UTR region. (Figures 7 and 8). Surprisingly, hu-UNC-53/1 identified also the cDNA clones gbH09036 and gbAA049124 homologous to the predicted coiled coil region in Ce-UNC-53 hu-UNC-53/1, and furthermore identified a third weakly homologous EST gbR21023. The inserts of gbH09036, gbAA049124 and gbR21023 were obtained from the Merck consortium and sequenced.

gbAA049124 is >95% identical to Hu-UNC-53/1 over 604 available amino acids (fig. 10) and is the mouse orthologue of Hu-UNC-53/1. The insert in gbH09036 is clearly homologous to hu-UNC-53/1 but derived from a different locus. We therefore name the gene identified by gbAA049124 Mu-UNC-53/1 and the gene identified by gbH09036 Hu-UNC-53/2. (Figure 11).

5 domains of high similarity mark the *unc-53* gene family

Ce-UNC-53 and the here-identified vertebrate homologues form a unique novel protein family, that is distant from the remainder of the proteins in the public domain. Alignment of the predicted open reading frames shows that Hu-UNC-53/1 and Hu-UNC-53/2 are equidistant from *Ce*-UNC-53. The highest homology is found in the carboxyterminal amino acids of *Ce*-UNC-53 region. The presence of a conserved GXXGKS/T box suggests a nucleotide binding function. However, this domain as a whole does not belong to a class of known nucleotide binding proteins.

The similarity amongst the presently known sequence of the UNC-53 family of proteins is highest in 5 blocks over most of the available sequence (959 amino-acids) and a firther block identified in Figure 12a. These blocks can be assigned signature sequences as displayed in figure 28 or can be assigned weight matrices based on the alignment between the different family members. By using truncated constructs of Ce-unc-53, the functional relevance of these domains has been addressed.

Hu-UNC53/1 and Hu-UNC-53/2 are complex transcription units.

1. A cancer cell line RNA blots probed with HU-Unc53/1.

A Northern blot of poly-A+RNA from several cancer cell lines (Melanoma G361, Lung Cancer A549, Colorectal Adenocarcinoma SW480, Burkitt Lymphoma DRajii, Leukemia Molt4, Lymphoblastic Leukemia K562,

HeLa S3 and Promyelocytic Leukemia HL60) was probed using the whole insert of pHH3b. No or weak expression was detected in the Burkitt Lymphoma DRajii, the Leukemia Molt4 and the Promyelocytic Leukemia HL60 cell lines. Five different transcripts are detected in the remaining cancer cell lines: transcripts 1 and 2 are larger than 9.5kb, transcripts 3 and 4 are 6 to 7 kb and the fifth transcript is around 6 kb. Transcripts 1 and 2 are present in all experssing cell lines. Transcripts 3 and 4 are restricted to Melanoma G361, Lung Cancer A549 and Colorectal Adenocarcinoma SW480 and are the predominant transcripts in Melanoma G361 and Colorectal Adenocarcinoma SW480. Transcript 5 is restricted to Lymphoblastic Leukemia K562 and HeLa S3 and is predominant in HeLa S3.

2. Cancer cell lines RNA blots probed with HU-UNC-53/2.

A similar set of cancer cell line Northern blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-aaacaccagtgagtcc. HU-UNC-53/2 is expressed in Melanoma G361, Colorectal Adenocarcinoma SW480, Lymphoblastic Leukemia K562 and HeLa S3. No expression was detected in Lung Cancer A549, Burkitt Lymphoma DRajii, Leukemia Molt4 and promyelocytic leukemia HL60. Interestingly only 2 transcript sizes were detected of around 7 kb expressed in Lymphoblastic Leukemia K562 and HeLa S3 and a transcript of >9.5 kb in Melanoma G361 and Colorectal Adenocarcinoma SW480.

3. Normal Human tissue probed with HU-Unc53/1.

A Northern blot of poly-A+RNA from normal

human tissue was probed using the whole insert of phage HH3b. Expression levels are low in all tissues with the highest level in heart and placenta, several fold lower levels in brain and testis, even lower levels in skeletal muscle, pancreas, thymus, colon, small intestine, ovary and prostate. Expression in peripheral blood leukocyte, lung, liver, kidney, spleenis barely detectable.

4. Normal Human tissue probed with Hu-unc53/2.

A similar set of blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-aaacaccagtgagtcc. Expression levels are low in all tissues with the highest level in kidney, lower levels in heart, placenta, lung, skeletal muscle and pancreas. Expression is barely detectable in brain and liver.

The hu-UNC53/1 and hu-UNC-53/2 homologues are clearly highly regulated genes, showing a strong tissue specificity and, probably, additional mechanisms of regulation (ie differential splicing of different promoters). The different proteins derived from RNA's identified by probe hh15 presumably share the carboxyterminal nucleotide binding domain. Ce-UNC-53 was shown to be a complex genetic locus and complex transcription unit. The different transcripts are thought to be a mechanism to assure the necessary specificity and functional diversity of this signal transduction pathway, with respect to different signals and receptors, different tissues and different directions of migration. The occurance of a new transcript or the observed changes in expression levels in the cancer cell line blot suggests a role

for hu-UNC-53/1 and hu-UNC-53/2 in the establishment or maintenance of the transformed state of those cells.

Phenotypic changes in cells transfected with the Nucleotide Binding Domain of Ce-UNC-53/1 and Hu-UNC-53/1

Ectopic expression of full length Ce-UNC-53 in C. elegans, murine neuroblastoma cells or human MCF-7 breast-carcinoma cells, has been found to lead to increased filopodia outgrowth and increased motility (unpublished). The structure of Ce-UNC-53 protein is reminiscent of that of large kinases or dynamin where a catalytic domain is postively or negatively regulated by domains that interface with signal transduction pathways for example (by by GRB2 binding, phosphorylation or the like). The inventors therefore decided to test whether the nucleotide domain by itself is capable of inducing the observed changes in the microfilament cytoskeleton and motile or ruffling behaviour.

cDNA fragments coding for the nucleotide binding domains of Ce-UNC-53 and Hu-UNC-53/1 were cloned in mammalian expression vectors with the CMV promoter (see experimental procedures).

To be able to detect expression from pCB201 (Fig. 13), an N-terminal his and a T7 epitope tag were fused in frame with the hu-UNC-53/1 cDNA hh15. pCDU3 contains a larger fragment of Ce-UNC-53 and starts just before the conserved "VIELKIEL" domain (Fig. 12).

The empty pcDNA3 vector or pCDNA3.1-His-LacZ, a mammalian expression vector for E. coli Beta-galactosidase, was used as a control vector (mock transfection). The differences between mock and

transfected N4 and MCF-7 clones were analysed using phase-contrast and Nomarski microscopy coupled with time lapse analysis, phagokinesis and immunocytochemical characterisation of the F-actin.

Phenotypic changes in mouse N4 neuroblastoma cells

N4 neuroblastoma cells were stably transfected with control construct pCDNA3.1 and the C. elegans UNC-53 constructs pTB72, pCDU2, pCDU3 and pCDU4. The population of clones transfected with the empty expression vector were homogeneous and similar to wild type N4 cells. In contrast thereto, 1/4 to 50% of the clones transfected with pTB72, pCDU2, pCDU3 and pCDU4 (see experimental procedures and Figs. 1,17,15 and 16 respectively) had distinct phenotypes:

1. Wild type or N4 cells transfected with pcDNA3, designated as mock transfection show a central cell body, with extensions, designated as neurite outgrowths. Less than 5% of the population have lamellae. When present, they are generally situated on the cell body and on the opposite site of the neurite extensions (figure 22a). The lamellae show a radial actin spike pattern. Limited branching of the actin fibres is observed in wild type or pcDNA3 transfected N4 cells. Side branches are smaller and can be clearly distinguished from the main actin branch (figure 23a).

2. N4 cells, stably transfected with pCDU4, harbouring the homology block E, show an overall morphology which is similar to that of wild type N4's (a cell body with neurite outgrowth). They exhibit however an increased frequency and level of lamellae formation (figure 22b). These lamellae, which contain

F-actin microspikes are found on both the cell body and the neurite outgrowth (figure 23b). Wild type N4 cells, in contrast thereto, rarely exhibit lamellae on the neurite outgrowths.

3. N4 cells, stably tranfected with pCDU3, encoding for homology blocks C, D and E, show an even higher level of lamellae formation labelled with TRITC-phalloidin, the cells appear surrounded with F-actin fibres, consisting of bundles of F-actin microspikes (figure 23c). The presence of these lamellae has completely modified the general appearance of the cells. They appear flatter and in 90% of the population, it is not possible to distinguish between the cell body and the wide neurite as they flow gradually into one another (figure 22c). If wild-type-like thin neurite-like outgrowths are present, they are frequently numerous, branched and located all around the cell.

4. The overall morphology of N4 cells, stably transfected with pCDU2, encoding for homology blocks A, B, C, D, and E, resembles that of the wild type cells since, cell body and neurite outgrowth can be clearly distinguished. The pCDU2 transfected cells however show more neurite outgrowth, and these are long and very branched, especially at the end of the outgrowth. When neurite outgrowths of different cells make contact, increased branching can be observed, giving the appearance of a network (figure 22d). N4 cells, transfected with pCDU2, show bundles of long radial F-actin filaments (microspikes), which can be branched, especially apically. The space between the hand-shaped actin spikes is mostly filled in with actin, leading to small lamellae-like structures. Also the network-like branching between the cells shows both the bundled actin structures and the

lamellae-like fill-in features. These dense F-actin structures are sometimes seen on the cell body, which enhances the network-like appearance of the cells (figure 23d).

5. N4 cells, stably transfected with plasmid pTB72, encoding the full length C. elegans UNC53 protein, seem to have a more rigid structure than wild type cells, most clearly seen as spindle-like and triangle-like cells. The corners of these cells show an increased level of hand-like lamellae structures. This specific phenotype is best seen when the cells are grown at low density (figure 22e, Fig. 23e).

Phenotypic changes in human breast carcinoma MCF-7 cells

MCF-7 cells were stably transfected with the pTB72, pCDU2, pCDU3, pCDU4 and pCB201. The population of clones transfected with the LacZ-expression vector were homogeneous and similar to wild type MCF-7 cells. In contrast thereto, ~30-50% of the clones transfected with pTB72, pCDU2, pCDU3, pCDU4 and pCB201 had distinct phenotypes which were analysed as above for the N4 cells:

1. Wild type and mock (pcDNA3) transfected MCF-7 cells are heteromorph. In general they are round cells or clusters of cells surrounded by lamellae. Bulges, similar to thick filopodia, can be observed (figure 19a). When the cells are stained with FITC- or TRITC coupled phalloidin, F-actin actin stress fibres can be observed, often in rings surrounding the cell body (figure 20a & 21a). When cells are round up like this actin is present at the edge of the cell body. Less than 10% of the cells display filopodia

filled with radial F-actin microspikes. In time-lapse analysis the cells are highly quiescent with limited ruffling at the edge of the cell.

2. MCF-7 cells transfected with pCDU4, encoding for homology block E, show two major phenotypic differences compared to the wild type cells. These cells are more flat and have more extended lamellipodia leading to a pancake-like appearance. Some clones show more filopodia than wild type (figure 19b). Radially organised F-actin fibres can clearly be observerd in the lamellae surrounding the cells. These stress fibres resemble the wild-type structures, but have a more radial than circular orientation. In the filopodia, one can observe an increase of apparently unorganised, bundles of actin patches (figure 21b).

3. MCF-7 cells, stably transfected with pCDU3, encoding the homology blocks C, D, and E, shows a strikingly different and constant morphology. The cells appear smaller than wild type because they are more rounded up. All the cells have more filopodia, surrounding the cell body (figure 19c). Morphologically these filopodia have the same "hand-like" appearance as those observed in N4 neuroblastoma cells. Such filopodia are hardly ever observed in mock transfected MCF-7 cells. These filopodia are filled with F-actin fibres. Compared to wild type cells, fine actin stress fibres are decreased (figure 21c). In time-lapse analysis single cells as well as clusters of cells can be seen to ruffle much more dynamically than single or clusters of wild type cells. The "half-life" of a filopodia outgrowth on the cell surface is much shorter in transfected cells and the numbers of filopodia present at any time higher.

4. Cells transfected with pCB201 (which is structurally similar to pCDU4 but human) has a phenotype that is nearly indistinguishable from that of cells transfected with pCDU3 except that the observed phenotype and ruffling activity and filopodia outgrowth is even higher than pCDU3 (figure 18).

5. The overall morphology of the MCF-7 cells transfected with pCDU2, which encodes the homology blocks A, B, C, D and E, resembles that of the pCDU3 transfected cells. The cells are more rounded up and show more filopodia than the wild type and mock transfected cells (figure 19d). The filopodia, which are all around the cell body tend to be longer, and show a difference in actin organisation. The small filopodia have the same actin bundles as seen in the pCDU3 transfected cells. In the longer filopodia, the actin bundles are more parallel, and radial to the cell body (figure 21d).

6. MCF-7 cells transfected stably with pTB72, encoding the full length UNC53 protein, are extremely rounded up, and tend to adhere more than wild type cells. The cells grow in clusters with sausage- or tube-like shapes. The presence of large extremely thin lamellae with a surface area of more than three times the central cell body forms a second morphological feature, unique for the pTB72 transfected MCF-7 cells (figure 19e). These sheets are difficult to observe under a phase contrast microscope, but are very clear when stained with phalloidin. The lamellae protrude from one side of a cell or group of cells and are filled with thin long criss-crossing actin fibres, different from "giant" wild type MCF-7 cells (figure 21e).

These experiments lead to the following set of conclusions: (Figure 47 summarises the data of the

domain swapping experiments in C. elegans unc-53)

1. Murine and human cells transfected with the Ce-UNC-53 or hu-UNC-53/1 domains show clear effects on the nature and dynamics of their motile behaviour as demonstrated by changes in the F-actin cytoskeleton (the increase in lamellipodia, hand-like filopodia and "hair-like" microspikes on the cell surface and the associated reduction of the "rings of F-actin" stress-fibres).

2. This effect is found in two cell types of different species and tissue origin: MCF-7 cells (human breast carcinoma cells of epithelial origin) and murine N4 neuroblastoma cells. pCB201, pCDU3 and pCDU4 induce in MCF-7 cells a type of filopodium which is frequent in wild type N4 cells but rare to absent in wild type MCF-7 cells, suggesting the activation by these constructs of motile behaviour which is "normal" in N4 cells but of an unusual type in MCF-7 cells. This indicates the activation of a specific downstream process as opposed to a disruption of an existing process. It is well known that some cell types prefer to migrate with filopodia and other cell types with lamellipodia.

3. Expression of pCB201, pCDU3 and pCDU4 gives qualitatively similar F-actin remodelling and increased filopodia and lamellipodia outgrowth. pCB201 and pCDU3 are however much more active in this process than pCDU4.

4. pCB201 is a much more potent activator of filopodia outgrowth than pCDU4, which is to be expected considering the large evolutionary distance between between C. elegans and vertebrates.

5. These experiments identify homology domain E (predicted nucleotide binding domain) of UNC-53 as the

"domain" that activates F-actin remodelling and filopodia/lamellipodia outgrowth. Progressive addition of the aminoterminal homology A,B,C,D lead to qualitative and quantitative modulation of the phenotype present in domain E.

6. Homology domains C and D (pCDU3) "enhance the basic activity present in homology domain E (pCDU4/pCB201).

7. Homology domains B and C (pCDU2) qualitatively modify the phenotype of domain E, leading morphologically different lamellipodia formation than pCDU3 transfected cells. It is thought that lamellipodia and filopodia formation are mediated by different signal transduction pathways requiring two related but different Ras-like G-proteins RAC for lamellipodia formation and CDC42 for filopodia formation.

8. pTB72 which includes homology domains A,B,C,D,E plus an additional 700 amino acids not yet identified isolated in the human members of the family confers a more localised filopodia outgrowth and a different morphology.

9. The expression levels of pTB72 (full length C. elegans UNC-53), pCDU3, pCDU4 and pCB201 are extremely low. The observed effect is therefore unlikely to be due to dominant negative effects (such as stoichiometric depletion of other cellular components) or structural changes in the actin cytoskeleton mediated by UNC-53 or its fragments.

The data point to a multi-domain organisation in UNC-53 whereby the aminoterminal domains exert positive (e.g. pCDU3) and negative (e.g. pCDU2) control on the activity of the domain E or are leading to novel activities or the localiation of the activity in the cell (pCDU2, pTB72). Our observation that the

nucleotide binding domains (NTB) of distantly related members of the UNC-53 family induce similar phenotypes, suggests a general role for this domain of the UNC-53 family.

CELLULAR ASSAYS TO IDENTIFY PHARMACOLOGICAL MODULATORS OF UNC-53 AND COMPONENTS OF THE UNC-53 PATHWAY

Mammalian and human cells transfected with plasmid constructs containing unc-53 sequence of either C. elegans or of human origin were observed to display obvious, specific and similar changes in comparison to mock or untransfected parent cells. These changes relate to the functioning of the cytoskeleton, in particular the F-actin cytoskeleton, to cell locomotion and directionally cell motility and reflect UNC-53 gene family members as capable of playing an integrator function in cell motility.

The cellular tools derived through transfection and derived functional assays with these cells not only enable characterisation of the motile phenotype typically observed after introduction of unc-53 genes, they also can be easily adapted to screen for pharmacological compounds that interfere with either (1) the expression of unc-53 gene family members, (2) the cellular functioning of unc-53 transgene(s) and of components in the unc-53 signal transduction pathway.

Two classes of pharmacological modulators are envisaged.

A first class are inhibitors of UNC-53s or the unc-53 pathway(s), which revert the described phenotypic changes induced by unc-53 transgenes or

aspects thereof. Such compounds are considered relevant leads to target diseases where unwanted directional motility of cells occurs such as metastatis, angiogenesis or inflammation.

Secondly, pharmacological stimulators are envisaged, such as compounds which induce - in non-transfected cells - phenotypes that induce or mimick (aspects of) the described `unc-53' phenotype. Such compounds may do so by inducing or upregulating expression levels of a known unc-53 gene or by activating endogenous (yet unidentified) members of the unc-53 gene family. The target application here are wound and tissue repair, in particular diseases such as neuronal regeneration and plasticity.

The nature of compounds envisaged can be small (organic) molecules, bio-molecules (such as peptides, sense or antisense (oligo-)nucleotides or chemical modifications thereof. Alternatively, compounds can be thought of as a series of plasmid nucleotide constructs containing gene sequences in a screen for novel unc-53-unrelated genes with a similar functional effect in the cell or genes related to the unc-53 gene family or novel members of the unc-53 gene family based on sequence similarity such as for example the genes in plasmids pTB72, pcDU3, pcDU4, pcDU2, pcB201, or modifications thereof such as for example epitope tagged, deletion, complementation or mutagenised nucleotide constructs.

The cellular assays envisaged in the claims have been exemplified for three cell lines: the human breast carcinoma cell line MCF-7, the mouse neuronal cell line N4 and the mouse fibroblast cell line NIH-3T3. Pharmacological assays are focused on quantification of endpoints in a high throughout screening mode. Many of the computer aids for

(semi-) automation are well known to the field and currently applied in the applicants labs. Given the subtlety of the phenotypes observed, primary focus was given to morphological assays that assess the phenotypes or aspects thereof.

The nucleotide binding domain of Hu-UNC-53/1 has transforming activity in NIH3T3 fibroblasts

Biochemical and genetic analysis suggest that UNC-53 functions in GRB-2 mediated signal transduction pathways controlling cell motility. The occurence of an altered hu-UNC53/1 mRNA pattern in cancer cell lines, moved us to investigate if whether hu-UNC53/1 plays a role in the transformed state of those cells.

Thereto, we tested the ability of the nucleotide binding domain of hu-UNC-53/1 and Ce-UNC-53 to transform NIH/3T3 cells. Construct pCB201 (hu-UNC-53), which induces ruffling behaviour and cell motility, were transfected into NIH3T3 cells. Positive controls included Myc and H-ras. Negative controls included empty vector adn Rac 1N17 and cdc42N17.

The cells that survived G418 selection were assayed for loss of contact inhibition (their ability to grow as foci). Positive controls included the combination of two well known oncogenes Myc and H-ras which were able to produce a high number of foci. The nucleotide binding domains of both Ce-UNC-53 and hu-UNC-53/1 are able to induce foci in this assay (Fig 24 & Table 1).

Table 1 Foci formation in NIH-3T3 cells stably transfected with pcB201

mock	pcB201
22	138
59	143

This suggests that the function of UNC-53 is not restricted to the activation of motility. UNC-53 may exert this additional function through the activation of as yet to be identified signal transduction pathways. Oncogenes frequently arise when a "controlling" domain and "activation" domain are separated though chromosomal rearrangements or integration of a part of a gene in the oncogenic virus. E.g. Erb Receptor tyrosine kinases, Ost a nucleotide exchange factor for Rac-1.

Hu-UNC-53/1 is localized to chromosome 1q31.1

Clone F226 (BACH-135 (014), Genome Systems, inc) was isolated from a human genomic BAC library using pCR231 as a probe and was confirmed by sequence analysis to be derived from the hu-UNC-53/1 locus. Purified DNA from clone F226 was labeled with digoxigenin dUTP by nick translation. Labeled probe was combined with sheared human DNA and hybridized to normal metaphase chromosomes derived from PHA stimulated peripheral blood lymphocytes in a solution containing 50% formamide, 10% dextransulfate and 2X SSC. Specific hybridization signals were detected by incubating the hybridized slides in fluoresceinated antidigoxigenin antibodies followed by counterstaining with DAPI. The initial experiment resulted in

specific labeling of the long arm of a group A chromosome. A second experiment was conducted in which an anonymous probe which was previously mapped to 1p34 and confirmed by cohybridization with a chromosome 1 centromere specific probe, was cohybridized with F226. The experiment resulted in the specific labeling of the long and short arms of chromosome 1. Measures of 10 specifically hybridized chromosomes 1 demonstrated that F226 is located at a position which is 52% of the distance from the heterochromatic-euchromatic boundary to the telomere of chromosome arm 1q, and that corresponds to band 1q31. At total of 80 metaphase cells were analyzed with 72 exhibiting specific labeling (Fig. 25).

Gains of DNA sequences in 1Q31 were found in more than 10% of primary bladder tumors (Genes Chromosom Cancer 12: 213-219 (1991)). A putative tumor suppressor gene located near the locus F13B on chromosome arm 1q31-q32 appears to be involved in the pathogenesis of medulloblastoma (Int. J. Cancer 67: 11-15 (1996)). Loss of heterozygosity in this region of chromosome I has been implicated in development of human hepatoblastoma. Partial trisomies of 1q31 were found in Ewing's Sarcoma cell lines isolated from patients Cancer Genet Cytogenet 12: 1-19 (1984).

HU-UNC-53/2 is localised to Chromosome 11p15.1

DNA from clone F329 from BAC for Hu-unc-53/2 was labeled with digoxigenin dUTP by nick translation and applied in the experimental settings used for FISH of Hu-unc53/1 with F226. The initial experiment with F329 resulted in the specific labeling of the mid short arm of a group C chromosome which was believed to be chromosome 11 on the base of size, morphology

and banding pattern. A second experiment was conducted in which a biotin labeled probe specific for the centromere of chromosome 11 (D11Z1) was cohybridised with clone F329. This experiment resulted in the specific labeling of the centromere in red and the mid short arm in green of chromosome 11. Measurements of 10 specifically labeled chromosomes 11 demonstrated that F329 is located at a position which is 65% of the distance from the centromere to the telomere of the chromosome 11p, an area which corresponds to band 11p15.1. A total of 80 metaphase cells were analysed with 72 exhibiting specific labeling.

Chromosome 11p15 is a region showing loss of heterozygosity (LOH) in a variety of human malignancies, primarily breast cancer (Ali et al., Science 238, 185-188 (1987); Winqvist et al., Cancer Res. 53, 4486-4488 (1993)) but also Wilms' tumor (Dowdy et al., Science 254, 293-295 (1991); Cowell et al., Br.J.Cancer 67, 1259-1261 (1993)), ovarian and testicular malignancies (Lothe et al., Genes Chromosomes Cancer 7, 96-101 (1993); Weitzel et al., Gynecol Oncol. 55, 245-252 (1994)) stomach cancer (Baffa et al., Cancer Res. 56, 268-272 (1996)), lung cancer (Ludwig et al., Int.J.Cancer 49, 661-665 (1991); Fong et al., Genes Chromosomes Cancer (1994)), infantile tumors of adrenal and liver (Byrne et al., Genes Chromosomes Cancer 8, 104-111 (1993)). Since LOH is believed to indicate inactivation of a tumor suppressor gene at the location where LOH occurs, the frequent LOH found at 11p15 in multiple human cancers suggests the presence of either a cluster of tumor suppressor genes or a single tumor suppressor in this region (Seizinger et al., Cytogenet. Cell genet. 58,

10080-10096 (1991)). Chromosome transfer studies have shown that chromosome 11 can suppress tumorigenicity of both human breast cancer (Negrini et al., Cancer Res.55, 3003-3007 (1995)) and Wilms' tumor cells (Dowdy et al., Science 254, 293-295 (1991)) and a gene (named HTS1 or ST5) that may be responsible for suppressing tumorigenicity in HeLa cells has been mapped to 11p15 (Lichy et al., Cell Growth Diff. 3, 541--548 (1992)). Abnormalities at 11p15 have also been identified in a variety of other cancers, including lung cancer (parental origin of 11p15 deletion) (Kondo et al., Oncogene 9, 3063-3065 (1994)), bladder cancer (Presti et al., Cancer Res. 51, 5405-5409 (1991)), myeloid leukemia (translocation) (Nakamura et al., Nat. Genet. 12, 154-158 (1996)), malignant astrocytomas and other primitive neuroectodermal tumors (deletions) (Fults et al., Genomics 14, 799-801 (1992)), rhabdomyosarcoma (Scrable et al., Nature 329, 645-647 (1987)) and hepatocellular carcinoma (Fujimori et al., Cancer Res. 51, 89-93 (1991); Wang et al., Cell Genet. 48, 72-78 (1988)). Recently a gene, TSG101, was cloned that is mutated in human breast cancer and deleted in uncultured primary human breast carcinomas (Li et al., Cell 88, 143-154 (1997)).

DIAGNOSTIC ASSAY USING THE DNA SEQUENCE OF HUMAN UNC-53S

The differential expression of human unc-53 transcripts in Northern blots of normal tissues versus transformed cell lines and the chromosomal locus of hu-unc-53/1 at 1q31 being a locus linked to three diseases, suggests the potential implication of hu-unc-53 genes in oncogenesis. By using the complete

DNA sequence of hu-unc-53/1 or /2 or fragments thereof in FISH, the potential involvement of these genes can be diagnosed in patients as exemplified in figure 26. Alike, the use of these hu-unc-53 sequences in diagnostic PCR assays can be used to determine overexpression of hu-unc-53s or fragments thereof.

Assay for microscopic phenotypic UNC-53 transfected MCF-7 cells

Mock and unc-53 transfected MCF-7 cells were seeded at low density in culture plates and allowed to adhere to the vessel. Light microscopic inspection at different time points either on live cells or after chemical fixation with Karnovsky's fixative revealed that in pcB201, MCF-7 transfected cultures a rounded shaped cell body with at their boundaries many filopodia. In contrast, mock or untransfected clones had a predominant `flat' phenotype - with little or no filopodia. Quantitative measurements confirmed the statistical significance of this shift in phenotype (table 2 below).

TABLE 2

Quantification of phenotypic changes in unc-53 transfected MCF-7 cells (*)

Transfection:	clone	no feet (**)	with feet (**)	fraction with feet
mock	e	34	8	0.19
		37	0	0
pcB201	2	17	92	0.84
		37	83	0.69
	16	27	62	0.70
		20	71	0.78
		13	85	0.87

(*) Clones were passaged thrice, frozen and stored.

Thawed cells were trypsinised at confluency, monodispersed, seeded in flasks and allowed to attach to substrate overnight to 48 hours. Cultures were fixed with Karnovsky fixative and inspected using phase contrast microscopy. In parallel experiments, resistance to geniticin was confirmed.

(**) values are expressed as cells per microscopic view.

Assay for ruffling and motile behaviour using automated time lapse

The dynamic changes in cells are well known in the field. Animations of e.g. actin ruffles in astrocytoma cells or od actin based cell motility in e.g. fibroblasts can be accessed (http://www.stc.cmu.edu/CLMIBhp/Imggallpg/Moviespg/actinruffle.mov) or (http://util.ucsf.edu/mitchi/Movies/migration.html) on the world wide web. The dynamic changes as a result of transfection with unc-53 can best be appreciated in time lapse video sequences. At high magnification, the \`filopodia' display arrays of microspikes with highly dynamic behaviour. A rough visual estimate suggests these phenomena to be at least 10-fold increased in pcB201 transfected cells relative to the mock-transfected MCF-7 cells. Animations of these clones in NIH-Image can be requested from author or applicant.

Time lapse video imaging probably is the most informative way to appreciate the unc-53-induced phenotype in MCF-7 and is amenable to high throughput screening in a pharmacological context. Time lapses compressing 5 minutes real time supply sufficient information to quantitate the intensity of the motile

behaviour of pcB201 transfected MCF-7 cells in e.g. 12 well plates. In addition, algorithms have been described in the field which can automatically compute the `motile area' of cells by comparing cells in two images appropriately spaced in time (van laerebeke etal., 1992, cytometry, 13, 1-8).

Assay for visualising unc-53-induced F-actin recruitment in MCF-7 cells

Cultures were chemically fixed, detergent extracted and fluorescently stained for F-actin (filamentous-actin) using fluorescently labeled phalloidin (Wieland et al., 1985, Int. J. Peptide & protein Res, 21, 3-10) which display in a more specific way the dramatic phenotypic changes to transfection with unc-53 transgenes. By using image capturing and analysis of the F-actin patterns, image analysis algorithms well known in the field can assess in an automated way, the f-actin filament positions, texture and distribution relative to the nuclear position or gravity point of the cells. Such algorithms are capable of discriminating phenotypic changes and thus also effects of pharmacological inhibitors of transgene-induced phenotypes as well as compound induced unc-53 like phenotypes in mock or untransfected cells.

Phagokinesis assay for unc-53-induced directionality and quantity of motility

The methods are described in the experimental section. Two cell populations with different motile behaviour in phagokinesis assays were observed. In table 3 below the fraction of mock and UNC-53

transfected MCF-7 cells that produced linear tracks in the phagokinesis assay are shown. In the mock transfected MCF-7 cells, 61% of the cells produce a round track (long and short axis less than 2-fold different) and 39% cells produced `linear' tracks (long and short axis more than 2-fold different). pcB201 transfected MCF-7 cells produced an increase of the fraction of cells displaying `linear' tracks to 50%. An increase in the fraction linear tracks was made for MCF-7 cells transfected with full sequence Ce-unc-53.

In addition, a significant increase of 50% in the median area of tracks of a culture vessel was observed in the pcB201 transfected MCF-7 cells versus mock transfected MCF-7 cells (Table 2). These observations suggest that pcB201 as well as pTB72 transfection into MCF-7 cells is capable of increasing *in situ* locomotion in Ce-UNC-53 MCF-7, e.g. by increasing spreading, ruffling, or other forms of non-directional motility in the `round' population. In addition the Ce-UNC-53 transgene in MCF-7 cells drives a fraction of the MCF-7 cells from non-directional motility (round tracks) into directional migration (linear tracks). Clone 2 thus provides a tool to analyse inhibitory or stimulating effects of pharmacological compounds on directionality or quantity of cell motility in relation to UNC-53.

Table 3. Analysis of motility in phagokinesis assays

Track morphology: fraction linear tracks

plasmid	clone	round	linear	l/r
Mock	z4	18	13	0.42
		17	11	0.39
		22	12	0.35
pCB201	Clone 2	16	9	0.36
		13	13	0.5
		7	8	0.53
		9	9	0.5

Track Size

Clone	average±SD	min	max	(N)
Z4	1626±188	1444	2011	(8)
Clone 2	2326±283	1989	2816	(8)

Assays for the localisation of unc-53 in the cell to microtubules or microtubule (+) plus ends

UNC-53s have been shown to reside on microtubules and preferentially on the microtubule (+)-ends of cells. This localisation represents an important feature of the UNC-53 family of proteins, which is rarely observed in other proteins. Absence of microtubule (+)-end binding in the protein APC following mutation has been implied in the role of APC in colon cancer (Smith et al., 1994, Cancer Res., 54, 3672). In analogy, it can be postulated that the proper functioning of UNC-53 also may depend on its specific localisation in the cell.

The methods used in the examples which prove the co-localisation with microtubules form a base for a series of assays for compounds which specifically

affect microtubule (+)-end binding of UNC-53s. To the skilled eye, the typical localisation of an UNC-53 protein on microtubules can be readily recognised and thus is sufficient for the interpretation of whether the treatment with a compound has affected the localisation of this UNC-53 (or a fragment thereof). Moreover, by combining the described methods (co-localisation) - well known to one skilled in the field and exemplified by the methods in the "experimental procedures" section - one can unequivocally confirm a compounds ability of abrogating (or promoting) microtubule and microtubule (+)-end binding.

Such an assay comprises contacting a cell culture of a cell line expressing an UNC-53 with a compound in the culture conditions proper for the said cell line, followed by an incubation and finally observation of the UNC-53 (or fragment) in situ by e.g. fluorescence microscopy (for GFP-chimeras) or by fixing the cell culture and performing an immunocytochemical staining for the UNC-53 (or fragment). For the co-localisation, methods such as immunocytochemistry for the microtubules of a cell or cell line combined with either immunocytochemistry for Ce-UNC-53 or Hu-UNC-53s or fluorescent detection GFP-UNC-53 chimeras are performed consecutively.

C.elegans-UNC-53 preferentially binds microtubule plus-ends or GTP-tubulin

Biochemical characterisation of UNC-53 has shown that UNC-53 binds the SH3 binding domains of SEM-5/GRB-2 and binds F-actin *in vitro*. GRB2 has been localised to the cortex of the cell and reported to be involved in the control of cell motility. To determine the *in vivo* subcellular localisation of Ce-

UNC-53, we transiently transfected COS, HepG2 and MCF7 cells with pTB72, an expression construct containing the full length Ce-unc-53 cDNA. This construct was previously shown to activate cell motility in N4 neuroblastoma and MCF7 cells. This construct gives high transient expression in COS cells, high to medium levels of expression in MCF7 cells and medium to low levels of expression in HepG2 cells. To visualise UNC-53, tubulin and F-actin, transfected cells were stained with various combinations of the anti-Ce-UNC-53 mab 16-48-2, rabbit anti-UNC-53 polyclonal, anti-tubulin mab YL1/2 and fluorescently labelled phalloidin.

At high levels of expression UNC-53 co-localises with the entire microtubule cytoskeleton, but at lower expression levels UNC-53 signal is restricted to the terminal regions of the microtubules at the plus ends. Very low levels of the expression yield a dot-like pattern in the vicinity of the cortex of the cell.

To map the MTB plus end domain of Ce-UNC53, we made two constructs pcDU2(figure 17) and pcDU3(figure 15) in which the aminotermus of Ce-UNC-53 is deleted. Proteins corresponding to these constructs are thought to be made *in vivo* from different unc-53 promoters. Transient transfections followed by immunolocalisation showed these proteins to be cytoplasmic. In stable transfections in N4 neuroblastoma cells and MCF7 cells they were shown to be no longer toxic to cells but cause highly increased activation of filopodia formation. We thus uncoupled (1) toxicity of Ce-UNC-53 from activation of motility and (2) microtubule binding from the activation of motility.

Analysis of the microtubule association of the C.elegans and Human 1 UNC53

To isolate the microtubule association domain of the C.elegans UNC53, N-terminal GFP fusions were made. C-terminal deletions on the fusion product revealed that the microtubule association was localised in the N-terminal half of the protein. A GFP fusion was also constructed with the Human1-UNC-53, to analyse the microtubule association properties of this protein. The association with microtubules was confirmed. A mouse anti sera was used to show the presence of native Unc-53 on microtubule plus ends of melanoma line G361. The epitope recognition of the antibody was confirmed by immunohistology experiments with mammalian cells, transiently expressed with pLM4, expression the GFP-hu1-UNC53 fusion protein.

Results

1. When transiently transfecting pTB72 in several cell lines C.elegans UNC-53 associates with microtubules and preferentially the plus-ends of the tubuline fibres. Transfection of plasmids pCDU3 and pCDU2 in N4 and MCF7 cell lines did not result in the observation of microtubule co-localisation. pCDU4 resulted in no staining using mab 16-48 antibody (LMBP Accession No. 1383CB) concluding that the epitope for this antibody is localised outside the fragment expressed by pCDU4.

It is possible that the microtubule associated domain is situated in the N-terminus of the protein. For this reason, we constructed an N-terminal GFP fusion with the full length C.elegans UNC-53 sequence, and various C-terminal deletion derivatives. These

fragments encode the N-terminal part of UNC-53 from 139 to 760 aa.

Furthermore, to analyse if the cloned fragment of hu1-unc53 also could be associated with microtubules, a plasmid encoding a GFP fusion with the hu1-Unc53 protein was constructed, and introduced into mammalian cells. A derivative of this construct was also constructed.

2.

a) Transient expression of C.elegans Unc-53 GFP fusion in N4 neuroblastoma lines

N4 cells where transiently transfected with pEGFP72, encoding a fusion protein of GFP and full length C.elegans unc-53 sequence. On an inverted microscope, the fluorescence of the GFP molecule could be followed in living cells. Cells which expressed low to medium levels of the fusion molecule showed a normal morphology after 18h to 30h. In these cells the co-localisation of the GFP fusion protein with the microtubules could clearly be demonstrated (figure 38a). In cells which demonstrated a low but still distinct GFP fluorescence, specific microtubule plus-end association could be observed (figure 38b). Cells expressing high levels of the GFP fusion protein tend to round up, in such a way that the microfilaments are difficult to visualise. After 48h, almost no GFP expressing cells can be found. It has previously been observed in transient expression of Unc-53, using plasmid pTB72, that the protein is toxic for the cells. The transient transfection experiments with the pEGFP72 plasmid gives the same observation, indicating that at least two features of the Unc53 protein are conserved in the GFP fusion protein, being

the microtubule association and the toxicity of the protein.

The transfected cells were fixed with paraformaldehyde, and the tubuline was stained using antibody YL1/2 and antimouse-CY3 (Jackson Labs). Although a significant loss of GFP fluorescence was observed, one could clearly demonstrate that the filaments observed with the GFP fluorescence co-localise with the microtubules staining (figure 39).

Putative Assay

Mammalian cells, in this case N4, were transfected with a lipofecting agent (lipofectAMINE) while in suspension, not being attached to a surface. After transfecting those cells with pEGFP72, the transfected cell suspension could be diluted in 24- and/or 96-well plates, enabling them to attach ot the surface. Each well may contain a different compound of the collection to screen. After 24h, plates could be automatically screened for fluorescence levels. Wells containing a compound that abolish the toxicity of the GFP-C.elegans UNC-53 fusion protein will give high levels of fluorescence. Compounds having no effect on the fusion product will give no or only low levels of fluorescence.

b) Transient expression of the truncated GFP-C.elegans UNC-53 fusion proteins.

To assay if the microtubuline association did occur in the N-terminal part of the C.elegans Unc-53 protein, various C-terminal deletions were constructed.

Transfection of pEGFPsma and pEGFPecl coding

for 760 AA and 670 of the N-terminal part of C.elegans UNC-53 in fusion with GFP, resulted in microtubuline association, as could be visualised in living cells. The association with the microtubules is less abundant than observed when expressing the full length C.elegans UNC-53 protein, but fibres could clearly be observed (figures 40a and 41a). More background fluorescence is seen. This could be due to a lesser association to the microtubules or to a instability of the fusion protein. The association with microtubules could not be observed after fixing the cells with paraformaldehyde nor with methanol fixation, giving an extra indication for the weak association with the microtubule network of these proteins or potential instability of the fusion protein. At low expression levels the association of the GFP fusion protein with the centrosomes could clearly be detected (Figures 40b and 41b). Centrosomes are the location in the cell with the highest microtubule concentration.

No plus-end associations could be observed with the deletion constructs, even when cells where expressing low levels of the GFP fusion proteins. In the case of very low expressions, the centrosomes could clearly be detected.

When transfecting N4 cells with pEGFPsac or pEFPXba, coding for 139 aa and 256 aa of the N-terminal part of C.elegans UNC-53 in fusion with GFP, no microtubule association could be observed. This indicates that at least 670 aa of the N-terminus of the C.elegans UNC-53 is needed to have microtubule association (figures 42a and 42b).

c) Transient expression of the GFP-hu-UNC-53/1

fusion proteins and a deletion derivative.

Plasmid pLM4 was transiently transfected into N4 neuroblastoma cells, and GFP fluorescence was observed in living cells. GFP fluorescence of the available sequence of hu1-UNC-53 in fusion with GFP was localised at the microtubule level. Moreover, at lower expression levels, both the centrosomes, and specific plus-end association could be observed. As has been observed with the C.elegans UNC-53 derivatives in fusion with GFP, expressed by the plasmids pEGFPsma and pEGFPec1, the GFP association seems to be less tight as was observed by the full length C.elegans UNC-53 fragment in fusion with GFP. The observed instability of the fusion protein can be due to a lesser association to microtubules, or to a degradation of the fusion protein (figure 43).

d) Immunofluorescence on melanoma line G361, and on neuroblastoma line N4 transiently transfected with pLM4.

Introduction

Northern experiments show that the melanoma cancer line G361 expressed abundantly both the Human1 and Human2 homologue of C.elegans UNC-53. To test if the proteins could be localised in this cell line, a collection of mouse sera was tested on this cell line. To verify if the observation was due to a hu-UNC-53 recognition, and not to an artifact, a positive sera was applied to N4 cells transiently transfected with pLM4, expressing the GFP-hu1-Unc fusion.

result

a serum, designated 28.1 from a mouse previously injected with peptide (DNRTLPKKGLYRY) a conserved sequence of the UNC-53 family was used for a immunolocalisation experiment on G361 cells fixed with paraformaldehyde. Antimouse-cy3 was applied as second antibody. Association with microtubule plus-end could clearly be observed. Moreover, in cells showing directional movement, observed as growth cones extensions, abundant staining can be seen in the tip of the growth cone (figure 45). To test whether the recognition of the microtubule associated protein was identical to the Hu1-UNC-53 protein, N4 cells were transiently transfected with plasmid pLM4 and consequently fixed with paraformaldehyde and stained with serum 28.1. Only cells that were transfected showed staining with 28.1, indicating that the antibody of 28.1 recognised the Hu1-UNC-53-GFP fusion protein (figure 46). This confirms that the staining of the microtubule plus-ends in the growth cones of G361 by serum 28.1 is due to a recognition of at least the Human1 and/or the Human2 homologue. It is concluded that the overexpression of the human homologue of C.elegans UNC-53 in the melanoma cancerline G361 is located on the microtubule plus-ends.

Conclusions

a) - GFP-C.elegans UNC-53 fusion protein expressed by pEGFP72 shows Unc53 activity

b) - GFP-C.elegans UNC-53 fusion protein expressed by pEGFP72 shows microtubule association

c) - GFP-C.elegans UNC-53 fusion protein

expressed by pEGFP72 shows microtubule plus-end association

c) - GFP-C.elegans UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPsma and pEGFPecl show microtubule association.

d) - GFP-C.elegans-UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPsma and pEGFPecl no not show microtubule plus-end association

e) - GFP-C.elegans UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPxba and pEGFPsac no not show microtubule associations.

f) - GFP-hu1-UNC-53 fusion protein expressed by plasmid pLM4 shows microtubule association.

g) - GFP-hu1-UNC-53 fusion protein expressed by plasmid pLM4 shows microtubule plus end association.

i) - serum 28.1 recognises the Hu1-UNC-53-GFP fusion protein as expressed by plasmid pLM4 in transiently transfected Neuroblastoma cells N4.

j) - the expressed human homologue of C.elegans.-UNC-53 in melanoma line (being at least hu1-Unc-53) is associated with the microtubule plus-ends.

EXPERIMENTAL PROCEDURES

Materials

The oligonucleotides used in the PCR-RACE experiments were synthesised by Eurogentee (Belgium). Radioactive compounds were obtained from Amersham. The pCDNA3.1 eukaryotic expression vectors, human lGT10 cDNA libraries, marathon-RACE cDNAS, human, Northern blots and the T7-tag monoclonal antibody were purchased from Invitrogen. N4, MCF7 and NIH 3T3 cells were retrieved from the Janssen Research cell bank.

PCR-RACE conditions

1. A quick screen human cDNA library panel was used to amplify EST clone gb..R41071. The primers used were ESTfw 5'-AATGGCTTCCTGGTTACCTGAG-3' and ESTrv 5'-CAAGTCAGCACCCCGAAGCAGCTCT-3'. Human genomic DNA was used also as template (100ng/reaction). The amplification conditions were as follows: 1 min at 94°C, 30 sec at 55°C, 30 sec at 72°c, then 35 more times and a final extension of 20 min at 72°C. This PCT fragment was cloned in vector pCR2.1. The resulting plasmid was designed pCR231.

A human heart clone was also produced by RACE-PCR from a human heart Marathon cDNA using the following conditions; 1 min at 94°C, 30 sec at 70°C, 3 min 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72 c KlenTaq DNA Polymerase was purchased from Invitrogen.

For the mouse homologue, total RNA was obtained from N4 murine cells as described. A first strand cDNA was synthesized from 2 μgr of RNA using Ready To-Go cDNA kit (Pharmacia) The primers used were M-ESTfw 5'CCTCTGTGGGCACCGAGGTCACC--3'. The amplification conditions were as follows: 1 min at 94°C, 30 sec at 58°C, 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72°C. All the amplifications product were subcloned in pCRII-1 and several independent clones were analyzed by sequence.

2. Screening of Human Heart/Colorectal Adenocarcinom a cDNA library

A human heart cDNA library and a human colorectal adenocarcinoma cDNA library were screened using pCR231bp as probe by the standard plaque hybridization

method. The screening produced several positive clones in each library called respectively λHH3, λHH4, λHH15, λCAD14 and λCAD27. The positive phages were purified by two additional rounds of plaque screening and were then amplified.

3. 5' extension using PCR

Three primers with homology to the 5' end of clone λHH3b were made: HU53rv1 (5'-cct-ggg-act-gaa-gct-ggt-acc-tga-gcc-3'), HU53rv2 (5'-ttg-gga-aga-gtg-ttc-cga-tcc-cgc-tg-3') and HU53rv3 (5'gtt-gcc-cag-ctc-tgg-ggc-ttc-cac-tcc-3') and used together with λgt10rv primer (5'-gag-gtg-gct-tat-gag-tat-ttc-ttc-cag-ggt-a-3') in three nested PCR reactions on a cDNA amplified library from Human Heart (Clontech). The reaction mixes contained 25pmol of each primer, 1 mM of each dNTP, 1 μlKlenTaq Polymerase Mix (50x) and 0.1 ng DNA. The cycling parameters for the first PCR were: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 51°C and 3 min at 72°C and a final extension of 10 min at 72°C, using HU53rv1 and λgt10rv as primers. 0.4 μl of this primary PCR product was amplified using HU53rv2 and λgt10rv as nested primers with the following parameters: 3 min at 94°C, 38 cycles of 1 min at 94°C, 1 min at 52°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. The second nested PCR reaction was performed on 0.4 μl of a 1/50 diluted purified 2.4 kb fragment using HU53rv3 and λgt10rv as primers: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 56°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. A 774 kb amplification product was subcloned in pCR2.1, resulting in plasmid pCB210-14. The clone fragment was analyzed by sequencing. This fragment

extends 699 bp in 5' direction (see fig 9).

4. 5' extension using PCR

Primer HU53rv4 (5'-ccc-tgc-ttg-gtg-ctg-agg-aga-ctg-g-3') was designed on the 5' end of clone pCB210-14 and was used together with λgt10rv to amplify a fragment of the Human Heart cDNA library with the following parameters: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 60°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. A 887 bp fragment was subcloned in pCR2.1, resulting in plasmid pCB212. The clone fragment was analyzed by sequencing. This fragment extends a further 767 bp in 5' direction (see fig 9).

5. Human Heart Library screening using the 0.8 kb insert of pCB212 as probe

The EcoRI digested and purified clone pCB212 was used as probe to screen the Human Heart cDNA library (Clontech) using standard plaque hybridization method. The positive phages were purified by two additional rounds of plaque screening. The insert of the λDNA (produced using Qiagen Lambda Kit) was analyzed by sequencing. This pHH14-3 resulted in a 2663 bp fragment overlapping pCB212, pCB210-14 and the 3' end (434 bp) of λHH3b and in a 761 bp 5' extension (see fig 9).

<u>3' and 5' extension of HU-Unc53/2 from EST46037</u>

<u>WashU-Merck EST 46037</u>

Transformed cells carrying the EST 46037 sequence were ordered from Research Genetics. Plasmid DNA was

isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

3' extension of EST 46037 by RACE

Marathon-Ready cDNAs (Clontech) are premade "libraries" of adaptor-ligated double-stranded cDNA ready for use as templates in RACE experiments. Five ml Marathon-Ready cDNA was used as template in a regular 50ml RACE. The RACE mixture contained 1x KlenTaq PCR buffer, 0.2 mM of each dNTP, 1x advantage KlenTaq polymerase mix (Clontech), 0.15 mM AP1 adaptor primer and 0.15 mM RACE gene specific primer. The amplification conditions were as follows :

94°C for 1 min, 5 cycles of 94°C for 30 s and 72°C for 4 min, 5 cycles of 94°C for 30s and 70°C for 4 min, 25 cycles of 94°C for 30 s and 68°C for 4 min. One-hundred-fold diluted RACE product was used as a template in a nested PCR with AP2 adaptor and gene specific nested PCR primers. Specific nested PCR fragments were cloned into pCRГ2.1 (TA cloning kit, Invitrogen) and the sequences of the inserts were determined.

gene specific primer (EST46037-F1) 5'AGTGAGAACAATGCTGTGGACATGC nested gene specific primer (ES46037-F2) 5'CTGCTCAACTGCAAGTACCACAAATGC Marathon cDNA library : human placenta

WashU-Merck EST 923793

Transformed cells carrying the EST 923793 sequence were ordered from Research Genetics. Plasmid DNA was isolated using standard protocols (Qiagen

plasmid DNA isolation kit), the sequence of the insert was determined.

RACE fragments 1.4 and 3.7, 5' extension of EST46037

Method as described previously. Gene specific primer (EST46037-R1) 5'ACTGCCTTGAGACTCTGACTTCAGC nested gene specific primer (ES46037-R2) 5'TGGGCAGAACTGAGAGCTTCTAAGC Marathon cDNA library : human placenta

RACE fragments B2.1, D2.1, H2.1; 5' extension

Method as described previously:gene specific primer (97010709) 5'ATTCTTTTGCATCTTCTTGCGTGCG nested gene specific primer (97010708) 5'ACCTGAGTCCTTTCTTAGGCAAAGTGTTCC Marathon cDNA library : human placenta (fragment B2.1)

human HeLa S3 (fragment D2.1) human colorectal adenocarcinoma SW480 (fragment H2.1)

PCR fragments E2.3, C2.3

EST 485068 is similar to but not identical with the 5'end of HU-Unc53/1. A primer pair consisting of one 3' EST 485068 primer and one 5' HU-Unc53/2 primer were used to PCR amplify those fragments. lgt10 human placenta Quick screen library (fragment C2.3) or Marathon cDNA from human HeLa S3 (fragment E2.3) were used as templates in a PCR. A 50 ml reaction mix contained 1xPCR II buffer (Perkin-Elmer), 1.5 mM MgCl2, 0.2 mM of each dNTP, 0.15 mM forward and reverse primer, 2.5 U AmpliTaq Gold (Perkin-Elmer)

and 1 ml template. The cycling parameters were 5 minutes at 95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 65°C and 2 minutes at 72°C. The PCR products were sliced out from an agarose gel and purified using a gel extraction kit (Qiagen), one ml hereof was used in a second round PCR using the same conditions as above. The PCR products were purified (Qiagen PCR purification kit) and direct sequenced.

primers :
(97010709) 5'ATTCTTTTGCATCTTCTTGCGTGCG
(97012802) 5' CGCTCCCCATCAGATGCAGGCCGG

PCR fragment E1.3-3

EST 01222 is homologous but not identical with the 5'end of HU-Unc53/1. A primer pair consisting of one 3' EST 01222 primer and one 5' HU-Unc53/2 primer were used to PCR amplify this fragments. Marathon cDNA from human HeLa S3 was used as template in a PCR. A 50 ml reaction mix contained 1xPCR II buffer (Perkin-Elmer), 1.5 mM MgCl2, 0.02 mM of each dNTP, 0.15 mM forward and reverse primer, 2.5 U AmpliTaq Gold (Perkin-Elmer) and 1 ml template. The cycling parameters were 5 minutes at 95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 65°C and 2 minutes at 72°C. The PCR products were sliced out from an agarose gel and purified using a gel extraction kit (Qiagen), one ml hereof was used in a second round PCR using the same conditions as above. The PCR products were analysed on an agarose gel, the fragment of interest was sliced out, purified (Qiagen PCR purification kit) and cloned into pCRΓ2.1. The sequence of the insert was determined.

RACE fragments A2.2-2, B2.1-4, D2.1-5; 5' extension

Method as described previously.
gene specific primer (97041701)

5'TATGCTACGGCCACTCATCTCCGTGG
nested gene specific primer (97041702)

5'TGTAACCTGAGTTCCCCTTAAACTGG
Marathon cDNA library :
human placenta (fragment A2.1-2)
human HeLa S3 (fragment B2.1-4)
human colorectal adenocarcinoma SW480 (fragment D2.1-5)

Translation-initiation splice variants, fragments D4.1-1, J4.1.4, G4.1.1, F4.1.2

Four different translation initiation slice variants were detected by 5'RACE.

Method as described previously.
gene specific primer (97080803)

5'TCGGTTGTTAGCAGTAGTTGACCCTCC
nested gene specific primer (97080804)

5'ACCTGAAAGTCTGGACTGCATTTCAGC
Marathon cDNA library : human colorectal adenocarcinoma SW480 (fragment D4.1-1) gene specific primer (97080801)

5'ACAACCTGGATAATCTGGGCCAGGAGG

nested gene specific primer (97080802)

5'TCTTGCTGGAGATCCTTGATGAGACGC
Marathon cDNA library :

human melanoma G361 (fragment J4.1.4)
human HeLa S3 (fragment G4.1.1)
human placenta (fragment F4.1.2)

DNA sequencing

PCR amplification products and cDNA clones were subcloned either into pBluescript vectors (Stratagene) or in PCR-IIa vector (Invitrogen) and sequenced either manually by the dideoxynucleotide chain termination method with modified T7 DNA polymerase (Sequenase, United States Biochemical) or automatically with an Applied Biosystems 373 DNA sequencer using the fluorescent terminator kit (Perkin Elmer).

RNA blots

A Human multiple tissue Northern (MTN-1, Clontech) containing in each lane 2 mg of poly A + RNA from eight different human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas) and a MTN-II human multiple tissue Northern, containing in each lane 2 mg of poly A + RNA from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral leukocyte, were hybrydized according to the manufacturer's instructions and washed out in 0.1xSSC:0.2% SDS at 55°C. Also from Clontech, a poly A + RNA blot from human cancer cell lines (melanoma G361, lung carcinoma A549, colorectal adenocarcinoma SW480, Burkitt's

lymphoma Raji Leukemia Molt 4, lymphoblastic leukemia K562, HeLa S3 and promyelocytic leukemia HL60) was tested.

Construction of plasmids

Plasmid pCDU2 (Figure 17) was constructed by cloning the 2.8 kb *ApaI-NarI* fragment from pTB72, the latter restriction site made blunt with klenow enzyme, into pcDNA3, digested with *EcoRV* and *Apal*. pCDU2 encodes for the homology blocks A, B, C, D and E. Plasmid pCDU3 (Figure 15) was constructed by cloning the 1.9 kb *Apal-NdeI* fragment from pTB72, the latter restriction site made blunt with Klenow enzyme, into pcDNA3, digested with *EcoRV* and *ApaI*, pCDU3 encodes for the homology blocks C, D and E. Plasmid pCDU4 (Figure 16) was constructed by cloning the 1.4 kb *ApaI-StyI* fragment from pTB72, the latter restriction site made blunt with Klenow, into pcDNA3 digested with *EcoRV* and *ApaI*. pCDU4 encodes for the homology block E.

Expression of a domain of the human UNC53 in eukaryotic cells

1. pCB201: Equivalent construct of human 1 homologue to expression construct pCDU4 of C. elegans unc-53 gene cloned in a eukaryotic His-tag, Xpress Ab tag expression vector.

A suitable Bam HI site was engineered on pHH15 open reading frame by amplification with hh15fw primer 5'AGAGCGGATCCATATGCCTCCTTGCCGTCAAGGTG-3' and M13rv primer (5'-cag-gaa-aca-gct-atg-ac-3'). The amplified fragment was then moved to pCDNA3.1.His-A-Vector

digested with BamH1 and EcoRI. This new plasmid called pCB201 (Figure 13) produces a cDNA which codes for a fusion protein consisting of a 49 amino acid aminoterminal fragment containing an His-tag and also a T7 epitope tag followed by amino acids 1255 to 1627 of the sequence of the human homologue. pCB201 was also checked by sequence and the n was used in stable transfection experiments carried out in N4, MCF7 and NIH3T3 cells.

2. pLM5: Equivalent construct of human 1 homologue to expression construct pCDU3 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector.

The phage HH3b was linearized using XhoI. A BamHI and XbaaI restriction site were created on the pHH3b open reading frame using U3-Bfw (5'-cca-cac-tag-ggg-atc-cat-gca-aat-gag-g-3') and U-rv (5'-caa-aag-tct-cta-gag-gag-gcc-agt-3') as primers. This amplified fragment was then moved to pBluescript KS, digested with BamHI and Xbal. Sequencing of this plasmid, named pCB300, showed an amino acid change from a serine to an asparagine due to a change from guanine to adenine on the position 4237 of the DNA sequence. This fault was repaired by cloning a 1418 bp fragment of pLM1 (see below) (using NarI and XbaI as enzymes) into pCB300 digested with the same enzymes. The phage HH3b fragment of this plasmid, named pLM6 (fig 53), was then removed using BamHI and XbaI, to pcDNA3.1/HisA digested with the same enzymes. This new plasmid, named pLM5 (fig 52), produces a cDNA which codes for a fusion protein consisting of a 49 amino acids aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by aminoacid 1069 to 1627 of the transcript of HU-Unc53/1. Plasmid pLM5 was

checked by sequencing and used on transient and stable transfection experiments carried out in N4 cells. The plasmid pLM1 was created using a PvuII and partial BamHI digested fragment of pHH14-3 and a BamHi and SpeI digested fragment of phage HH3b, cloned into pBluescript KS digested with SmaI and SpeI. The pLM1 contains the full transcript of HU-UNC-53/1 available at this moment (see fig 9).

3.pCB251: Equivalent construct of human 1 homologue to expression construct pCDU2 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector

The phage HH3b was linearized using XhoI. A BamHI and XbaI restriction site were created on the pHH3b open reading frame using U2fw (5'-aag-gga-tga-ttc-ggt-cag-gat-cct-tc-3') and U-rv (5'-caa-aag-tct-cta-gag-gag-gcc-agt-3') as primers. The amplified fragment was then moved to pCR2.1. This plasmid was named pCB250. The pHH3b fragment was removed from pCB250 using BamHI and XbaI and cloned in pcDNA3.1/HisC digested with the same enzymes. This plasmid, named pCB251 (figure 55), was checked by sequencing. pCB251 produces a cDNA which codes for a fusion protein consisting of a 49 amino acid aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by amino acids 828 to 1627 of the partial transcript of HU-Unc53/1. pCB251 was used on transient and stable transfection experiments carried out in N4 cells (see fig 56).

4. pLM3: the partial transcript of HU-Unc531 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector

pLM1 was digested with EcoRV and XbaI. This fragment was cloned in pcDNA3.1/HisB, digested with the same enzymes. pLM3 produces a cDNA which codes for a fusion protein consisting of a 49 aminoacid aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by amino acids 1 to 1627 of the transcript of HU-Unc53/1 available at this moment. pLM3 was used on transient and stable transfection experiments carried out in N4 cells.

5. pLM4: the partial transcript of HU-Unc53/1 cloned in an eukaryotic GFP expression vector

pLM1 was digested with ClaI and XbaI. This fragment was cloned in pEGFP-c1, digested with AccI and XbaI. This plasmid was named pLM4. This plasmid produces a cDNA which codes for a fusion protein consisting of GFP, followed by aminoacid 1 to 1627 of the transcript of HU-Unc53/1. pLM4 was used on transient and stable transfection experiments carried out in N4 cells (see figs 43 and 46).

__Stable transfection of MCF-7 cells__:

Cells were seeded at a density of $2x10^6$ cells in a 75 cm^2 flask using standard culture medium ((Dubecco's MEM, 450 mg/l glucose, 862 mg/l L-Alanyl-L-Glutamin, 110 mg/l Na-pyruvate; GibcoBRL) supplemented with 10% foetal calf serum (FCS; GibcoBRL), and 100 U/ml penicillin (GibcoBRL) and 100 μg/ml streptomycin). The culture was grown at 37°C in a 10% CO_2 atmosphere, to approximately 70% confluency (approximately 18 hours). The culture medium was removed and 10 ml MEM-HEPES (GibcoBRL) supplemented

with 10% FCS was added to the cells. The culture was further incubated for four hours at 37°C in standard sterile air. DNA-$CaCl_2$ was meanwhile prepared by mixing 30 μg DNA in 0.1 x TE (1 mM Tris. Hcl, pH 7.2, 0.1 mM EDTA, pH 8) and 0.1 ml 1.25 M $CaCl_2$/HEPES (1.25 M $CaCl_2$, 0.125 M HEPES; pH 7.05). 0.1 x TE was added to a final volume of 0.5 ml. The DNA-$CaCl_2$ was added drop by drop to 0.5 ml BS/HEPES (25 mM HEPES, 0.25 M NaCl, 0.01 M Kcl, 1.4 mM Na_2HPO_4, 0.01 M glucose, pH 7.05) while pipeting a sterile airflow through the latter solutions. The DNA-$Ca_3(PO_4)_2$ precipitate was then placed at 37°C for ten minutes. The DNA-$Ca_3(PO_4)_2$ precipitate was vortexed and added to the cells, together with 100 μl of a 0.01 M chloroquine (Sigma) stock in H_2O. After four hours of incubation at 37°C in sterile standard air, the medium was removed, and the cells were washed with PBS (GibcoBRL). 25 ml of medium was added and the cells where placed at 37°C in a 10% CO_2 atmosphere. After 48 hours of incubation, the cells were harvested, diluted and cultivated under selection (600 μg/ml G418 (Duchefa)) for two weeks prior to clone selection. Mock transfected MCF-7 were positive for the beta-galactosidase transgene. The stability of transfection in MCF-7 was assessed by passaging cells four times in the absence of Geneticin and then re-exposing them to the selector agent. In these experiments, unc-53 or mock transfected cells proliferated, whereas untransfected MCF-7 cells proliferated at a much slower rate.

<u>Stable transfection of N4 neuroblastoma cells</u>

Cells were seeded at a density of $2x10^6$ cells in a 25 cm^2 flask using standard culture medium ((MEM Rega 3; GibcoBRL) supplemented with 10% FCS, 0.14%

Na_2CO_3, 2 mM glutamine, 100 U/ml penicillin, and 100 μg streptomycin). The culture was grown overnight at 37°C in a 10% CO_2 atmosphere. Transfection mixture was prepared by adding 12 μg DNA in 600 μl optimem 1(GibcoBRL) to 36 μl LipofectAMINE (GibcoBRL) in 600 μl optimem 1. This was done by adding drop by drop the first solution to the second. The mixture was placed for 30 minutes at room temperature, after which 1.8 ml of optimem 1 was added. In the meanwhile the cell culture was washed twice with optimem 1, and the 3 ml of transfection mixture was added. The culture was placed at 37°C in sterile standard air. After four hours, 3 ml or normal culture medium was added and the culture was placed at 37°C under 10% of CO_2. 18 hours later, the culture was washed with PBS, and fresh normal culture medium was added. A further 24 hours later, the cells were harvested, diluted and cultured under selection (750 μg/ml G418) for two weeks prior to clone selection.

Fixation of cells for Immunofluorescence

Medium was removed from the 9 cm^2 wells containing the coverslips. A 4% solution of paraformaldehyde (Sigma) in PHEM (1 g/l glucose, 0.4 g/l Kcl, 8 g.1 NaCl, 0.06 g/l KH_2PO_4, 0.0475 g/l Na_2HPO_4, 0.35 g/l $NaHCO_3$, 1.51 g/l PIPES, 0.76 g/l EGTA, 0.19 g/l $MgCl_2$; pH 6) was added for 30 min at room temperature. The fixative was removed, and the coverslips were washed three times 10 minutes with PHEM. The coverslides were then placed in PHEM, containing 0.5% Triton-X100 (Serva) for 30 minutes, after which th slide was washed again for three times 10 minutes with PHEM. The coverslips were then placed

under PBS (0.14 M NaCl, 2.7 mM Kcl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) containing 0.2% Tween (Sigma) for at least one hour at 4°·C

Immunofluorescence staining

The coverslips were inverted on 35 µl of appropriately diluted antibody, being YL 1/2 for tubulin and/or mab 16-48-2 monoclonal or anti-UNC53 (gp48) polyclonal antibody for UNC53. The slides were placed at 4°C for at least 18 hours. Excess of primary antibody was then removed by washes of three times ten minutes in PBS-Tween. The slides were then treated with secondary antibody in the same way as for the primary antibody. F-actin was labelled by including TRITC- or FITC coupled phalloidine to the incubation buffer. The inverted slides on the secondary antibody were left at room temperature for approximately one hour. Slides were then washed again for three times ten minutes with PBS-Tween and once with PBS. The coverslips were mounted on slides with the medium described by Herzog et al. (Cell Biology: a laboratory handbook, 1994, Academic Press, 355-360). After at least two hours, slides were ready for analysis.

Time lapse analysis

Analysis of the behaviour and movement of growing cell cultures was done by placing a non-confluent culture under a phase contrast microscope equipped with a temperature controlled stage (37°C). Images were recorded using a CCD camera (COHU 4912) coupled to a SCION LG3 framegrabber in a Macintosh ppc 8100 running NIH image version 1.60. Images were recorded

at time intervals, varying from 15 sec to 1 min. for half an hour to two hours. Image enhancement and playback was done in NIH image.

Phagokinesis

A variety of cell types were shown to migrate over colloidal gold coated culture plastic or glass and displace or phagocytose the gold lawn on their way while locomoting. The track left bare is a qualitative and quantitative measure of cell motility and/or locomotion. The basic methods have been described in detail elsewhere (Albrechr-Buehler, 1977, Cell, 11: 395, Zetter, 1980; Nature, 285: 41; O'Keefe *et al.*, 1983; J. Invest. Dermatol., 85: 130). Culture plates were gelatin and gold coated as described by Albrecht-Buehler (1977). Unc-53 and mock transfected MCF-7 were seeded in plates at low density and allowed to adhere to the plate and to locomote overnight. Cells were chemically fixed to the plate, washed and air-dried. Images of the gold lawns were captured using automated videomicroscopy; composite images of the wells were generated and single-cell phagokinesis tracks were measured using a home-made routine in SCIL™ software.

C. elegans-UNC-53 preferentially binds microtubule plus ends or GTP-tubulin

1. Cloning of C.elegans cDNA in pEGFP-C1 and construction of C-terminal deletion derivatives.

a) Constructing a GFP-Unc53 N-terminal fusion:

A PCR experiment was performed under standard conditions, using pTB72 as template and cp17

(ata gcc aga tct acg tca aat gta gaa ttg) and cp18 (ttt aga aac cgc ggg tgg) as primers. The resulting 0.4 kb fragment, coding for the N-terminal fragment of C.elegans Unc 53 was cloned in vector pCR2.1 (original TA cloning kit, Invitrogen), resulting in plasmid pTA1718. The 0.4 kb fragment was isolated as a BglII-SacII fragment and cloned in pEGFP-C1 (Clonetech) digested with the same enzymes. The resulting plasmid was designated pEGFPsac (Figure 29). pEGFPsac encodes the N-terminal 13 aa of C.e.Unc53 in fusion with GFP.

b) Construction of a GFP-C.e. Unc53 full length fusion:

Plasmid pTB72 (shown in Figure 1) was digested with restriction enzymes SacII and ApaI. The resulting 4.5 kb cDNA fragment, encoding for the C-terminal fragment of C.elegans Unc53 was cloned in plasmid pEGFPsac (Figure 29), digested with the same enzymes, resulting in plasmid pEFP72 (Figure 30). Plasmid pEGFP encodes GFP in fusion with the full length C.e. Unc53.

c) Construction of N-terminal deletions of GFP-C.elegans UNC-53 fusion protein, other than pEGFPsac:

pEGFP72 was digested with SmaI. The resulting 7.0 kb fragment was religated and transformed in E.coli, resulting in plasmid pEGFPsma (Figure 31). This plasmid codes for the first 760 aa of the Ce-UNC-53 in fusion GFP.

pEGFP72 was digested with restriction enzymes Ecl136II and SmaI, the resulting plasmid after ligation and transformation in E.coli of the 6.7 kb fragment was designated pEGFPecl (Figure 32). This plasmid codes for the N-terminal 670 aa of the C.e. Unc53 in fusion with GFP. pEGFP72 was further digested with SmaI and XbaI. The latter site was made blunt with Klenow polymerase. The resulting fragment

of 5.4 kb was religated and transformed in E.coli. The resulting plasmid was designated pEGFxba (Figure 33). This plasmid codes for the N-terminal 256 aa of C.elegans Unc53 in fusion with GFP.

2. Constructing a hu1-UNC-53-GFP fusion, and a deletion derivative

The 5.4 kb hu1-unc53 fragment was isolated as ClaI-XbaI fragment from pLM1 (Figure 54), and cloned in pEGFP-C1 digested with AccI and XbaI. pEGFP-C1 was isolated from E.coli GM41 (Hfa H, dam-3, thi-1, rel-1). This makes the XbaI restriction site available for restriction digest. The resulting plasmid was designated pLM4 (Figure 34).

3. Visualisation of GFP fluorescence in N4 cells

N4 neuroblastoma lines where seeded in Lab Tek chambered coverglass (Nalge Nunc International) and transfected using lipofectAMINE (GibcoBRL). After 18 hours, the chambered coverglasses where placed on a inverted microscope, and GFP fluorescence could be visualised.

4. Staining GFP fusion expressing cells with antibodies

Transfection with the GFP fusion constructed was also performed on coverglasses in a 6-well plate. After paraformaldehyde or methanol-acetone fixation, cells could be stained for actin cytoskeleton with TRITC-phalloidine, for hu-unc53 with sera 28.1 and for tubuline with YL1/2 antibody. Visualisation was then

performed on a axioplan (Zeiss microscope).

Methods of Producing and Observing the Effects of A Chimeric unc-53 Gene

1. Definition of a promoter region in the unc-53 C.elegans gene:

The genomic region from the position 15621 to 18415 in the C.elegans unc-53 gene, called promoter A, was cloned and fused to the cDNA of the GFP gene (clone pA/GFP, or pNP10)(cf. fig.51). This construct is injected into wild type worms (N2). Transgenic line express GFP in different neurones: the two pairs of pioneering neurones PVP and PVQ, both BDU neurones, both ALN and PLN neurones, both PDE neurones, both PVM neurones, and 4 vulval cells. Expression begins in early embryogenesis, when the axons of those neurones grow out.

2. Mutant Phenotype in Unc-53(n152) alleles:

In wild type worms (N2), the two pairs of ALN and PLN neurones each send an axon in a straight line longitudinally from the tail to the head (see fig.50a). In unc-53(n152) alleles, the axons are shorter and often branch in a dorso ventral direction (see fig.50b). The neurones are visualised with the construct pA/GFP, injected in unc-53(n152) worms.

3. The minigene pA/unc-53 rescues the elongation defect of ALN and PLN neurones:

The promoter A from the C.elegans unc-53 gene was fused to the cDNA of the C.elegans unc-53 gene (clone pA/unc53, or pNP9). This construct was injected in unc-53(n152) mutant worms, together with the pA/GFP construct described above to visualise the ALN and PLN neurones. The elongation defects of those neurones in the unc-53 mutant are almost completely restored by the expression of the unc-53 cDNA express under the promoter A (see figs. 50 and 51b).

4. Domain swap between the C.elegans and human unc-53 gene:

To test whether the vertebrate and worm members of the unc-53 family are functionally equivalent, we tested the ability of the human gene to rescue the mutant phenotype in the worm. We replaced the carboxyterminal predicted nucleotide binding domain (NTPase) of the worm protein with the homologius fragment of the human 1 gene.

The clone pA/unc-53 was deleted of the C.elegans NTPase domain, from the HpaI site, position 29800 on the genomic of unc-53, and replaced by the equivalent domain of the human-1 gene (unc-53H1) (see fig. 51). The resulting clone is named pA/unc-53H1. When this clone is injected to unc-53(n152) mutants, the transgenic worms show a significant but incomplete rescue of the defect in the elongation of the ALN and PLN neurones (see fig. 51b). The axons are longer, often elongated until the region of the vulva in a straight line, without branching dorsally anymore. This result shows that a NTPase region of the human unc-53 homologue can functionally replace the NTPase region of the C.elegans worm.

The degree of rescue was analyzed quantitatively and summarized in Figure 51b:
The four strains compared are:
wt; un-53(n152); unc-53(n152),pA/unc-53;unc-53(152),pA/unc-53-H1.
The various phenotypes observed are brought together in three large classes:
<<wild type>> the axon is straight, unbranched and migrates into the head;
<<vulva>> the axon is straight, unbranched and stops in the vulva region;
<<mutant>> the axon is short, does not reach the vulva region and has collateral branches.
The figures are indicated as a percentage. The number of axons observed is indicated in the following column.

The data clearly show demonstrate conclude that the nematode/human chimera minigene pA/unc-53-H1 partly rescues the defects of the axonal migration of the ALN and PLN neurones and demonstrate conservation of function of this domain between man and worm. The transgenic lines provide a functional screening assay for the motility function of at least part of the human UNC-53 gene.

II. Materials and methods

1 - Cloning:

a) pAB/GFP (pNP3 - Figure 27)

The gene of GFP has been amplified by PCR with cpn3 oligo-nucleotides "acattaagcttcgtacgcttggagggtaccg" and Cpn5 "gaaaggatccgtacgataaggtattttgtgtcgg" on the plasmid pPD95.75(Figure 59) so as to be inserted at the 5'

position in fusion into the exon 12 of the unc-53 gene at a single restriction site SplI and contains its stop codon at 3' plus one polyadenilation site. The PCR amplification product is directed by HindIII and BamHI, sites which are contained respectively in the cpn3 and cpn5 oligonucleotides and sub-cloned in the pBS vector (clone pNP2). The GFP is then excised from the pNP2 clone at the site SplI and integrated into the X16 clone (Figure 60) originating from sub-cloning of the lambda phage S4 digested by Xhol. The X16 clone containing the genomic sequence of unc-53 from the position 16621 to the position 24891 cloned in the site XhoI of pBS.

b) pAB/unc-53 (pNP8 - Figure 35)

The promoter region AB of the X16 clone (between PstI and SplI) has been inserted in the clone pTB115 (Figure 58) in which the region between the sites PstI and SplI, containing the promoter of the gene mec-7 and the start of the gene unc-53, has been removed.

c) pA/GFP (pNP10 - Figure 56)

The promoter region A has come from the X16 clone between the sites PstI and NheI and integrated in the vector pPD95.75 containing the GFP in the sites PstI and XbaI.

d) pA/unc-53 (pNP9 - Figure 44)

The promoter region A has come from the X16 clone between the sites PstI and BstXI and is integrated into the clone pTB115 in which the region between the sites PstI and BstXI, containing the promoter of the gene mec-7 and the start of the gene unc-53, has been removed.

e) pA/unc-53 -H1 (pCB501 - Figure 57)

The clone pA/unc-53 (pNP9) has been deleted from the region 3' of the gene unc-53 of the nematode between the sites HpaI and NcoI. The 3' region of the H1unc-53 gene has been amplified by. PCR with the oligonucleotides U4Afw (5'-gca-cat-cgt-taa-cgg-gga-ctt-gaa-gc-3') and Urv (5'-caa-aag-tct-cta-gag-gcc-agt-3') and digested with HpaI and XbaI. After a filling stage with T4 polymerase, the ligation is effected with a complete end.

2-Injection

Conventional injection techniques are used (Fire A., 1986, Mello G, et al, 1991, journal Mello G. and Fire A., 1995). Young hermaphrodite adults are injected in their two syncytial gonads. The DNA used is prepared in standard manner (Qiagen) followed by precipitation with lithium chloride. After an extensive rinsing stage to eliminate all the salts, the DNA is resuspended in water. The injection solution contains the different DNAs at a concentration of 100 ng/ìl in an injection buffer:. The plasmid pRF4 containing the dominant allele su 1006 of the gene rol-6 (Kramer J. et al, 1990, Mello C. et al, 1991) is used as a transformation co-marker. The descendants of roller phenotype of the hermaphrodite injected are isolated. Approximately 10 % of these transformants will yield a stable strain, in which the different DNAs injected are associated to form a mini-chromosome which will segregate as unstable extrachromosomal arrays. All the transgenic strains obtained were tested by PCR for the presence of the DNA injected, using a specific primer of the vector and a primer in the gene (results not shown).

3. Microscopy

The nematodes are observed under a ZEISS Axioplan microscope provided with Nomarski lenses, with 40X Neofluar, 63X Plan-Apochromat, 100X Plan-Apochromat objective lenses. For fluorescence observation the luminous source is a mercury bulb. Different ZEISS filters are used:

- for observation under GFP fluorescence, FITC filter: blue excitation line at 588 nm, emission through a 515-565 nm band-pass filter;
- for observation of the antibody labelling with a secondary antibody coupled to the TRITC: excitation through a 546 nm band-pass filter, emission through a 590 nm long-pass filter.

The image acquisition is effected by means of a CCD camera and and NIH image program using a Machintosh computer. The images are processed using the Adobe Photoshop program.

Sequence Listing

The following sequences are referred to in the specification:

Sequence ID No 1 is an amino acid sequence of human homologue 1 of UNC-53 protein illustrated in Figure 9b.

Sequence ID No 2 is an amino acid sequence of human homologue 2 of UNC-53 protein illustrated in figure 11d.

Sequence ID No 3 is a nucleic acid sequence of the hu-1-unc-53 gene illustrated in Figure 9b.

Sequence ID No 4 is a nucleic acid sequence of the hu-2-unc-53 gene illustrated in Figure 11d.

Sequence ID No 5 is a nucleotide sequence of Phage Lamda Clone 3b deposited under Accession No LMBP 3595.illustrated in Figure 9.

Sequence ID No 6 is a nucleotide sequence of plasmid pLM1 deposited under Accession No LMBP 3762 and illustrated in fig 54.

Sequence ID No 7 is a nucleotide sequence of plasmid pLM4 deposited under Accession No 3763 and illustrated in fig 34.

Sequence ID No 8 is a nucleotide sequence of plasmid pEGFP72 deposited under LMBP Accession No 3764 and illustrated in fig 30.

Sequence ID No 9 is a nucleotide sequence of plasmid pCB501 deposited under Accession No 3765 and illustrated in fig 57.

Sequence ID No 10 is a nucleotide sequence of plasmid pCB201 deposited under Accession No. LMBP 3594.

SEQ ID No. 1

F .ino Acid Sequence of Hu-UNC-53/1 Protein

Page 1

GATATCTGCAGAATTCGGCTTCTTTGAGCAAGTTCAGCCTGGTTAAGTCCAAGCTGAATTCCGGGGAAAGCCGAGCCGGATCCCTCGAGGACCCTATGCG 100
CTATAGACGTCTTAAGCCGAAGAAACTCGTTCAAGTCGGACCAATTCAGGTTCGACTTAAGGCCCCTTTCGGCTCGGCCTAGGGAGCTCCTGGGATACGC

pCR2.1 linker — lambda gt10 primer EcoRI — suspect sequence linker? — pHH14-3

I S A E F G F F E Q V Q P G . V Q A E F R G K P S R I P R G P Y A

GAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGCGGCCGTGAGCGAAGATGGCAAATCGGACGACGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCT 200
CTCCAGTTCGGCGAGTCGTTCCGCGGACTTCGCCGGCACTCGCTTCTACCGTTTAGCCTGCTGCTCGACGAGAGGTCGTTCCGGTTCCGCGTTTTCTCGA

pHH14-3

E V K P L S K A P E A A V S E D G K S D D E L L S S K A K A Q K S

CTGGGCCTGTCCCCTCTGCCAAGGGCCAGGAGGAGCGCGCCTTCCTCAAGGTGGACCCCGAGCTGGTGGTGACCGTGCTGGGAGACCTGGAGCAGCTGCT 300
GACCCGGACAGGGGAGACGGTTCCCGGTCCTCCTCGCGCGGAAGGAGTTCCACCTGGGGCTCGACCACCACTGGCACGACCCTCTGGACCTCGTCGACGA

pHH14-3

S G P V P S A K G Q E E R A F L K V D P E L V V T V L G D L E Q L L

CTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGA 400
GAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCGGACGCT

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S L R

GGCTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTC 500
CCGAGGGTCCACTGAGTGTCGAGGGACCTCTACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAG

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

G S Q V T H S S L E M T C Y D S D D A N P R S V S S L S N R S S P

TGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTACAT 600
ACAGTACCGCGATACCGGTCAGGTCAGGCGCCGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGACCATGTA

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

L S W R Y G Q S S P R L Q A G D A P S V G G S C R S E G T P A W Y

Fig 9

Tuesday, 18 November 1997 10:33 Page 2
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GCACGGCGAACGGGCCCACTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCG 700
CGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGC

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L D S

GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCA 800
CTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTACTTTCGT

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G W D E S

GCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC 900
CGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACG

pHH14-3

pCB212

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P T A

TTCTCGCAGGAACTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCC 1000
AAGAGCGTCCTTGAGTTGTTATCACGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTCTTTCGG

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

S R R N S T I V L R T D S E K R S L A E S G L S W F S E S E E K A

CCTAAAAAACTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCATCCAAGG 1100
GGATTTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCC

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

P K K L E Y D S G S L K M E P G T S K W R R E R P E S C D D S S K

Tuesday, 18 November 1997 10:33 Fig 9

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence Page3

GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCACACAGC
CACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCG 1200

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGAT
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGGAGACTA 1300

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCCTGAGTGATGCTAAGAAGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCA
CGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT 1400

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC
GTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCTG 1500

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R K T

TAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGT
ATCGAATCTACAAAGGTTGTCACGTCTCGGTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCA 1600

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A K S S

Fig 9

Tuesday, 18 November 1997 10:33 Page 4

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

TCTATGAGCGTGACCGGCGGGCGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCA 1700
AGATACTCGCACTGGCCGCCCGCCCCACCTGGAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGG

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HUS3rv4

S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T P S

GACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGC 1800
CTGACTTCCTCGGATGGTTCCATCGGTCACCCGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCGTCACCG

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

R L K E P T K V A S G R T T P A P V N Q T D R E K E K A K A K A V A

CTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCAGAGCTG 1900
GAACCTGAGTCTGTTGTAGAGGAACTTCTCATAACCGAGGGGTCTTTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGTCTCGAC

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A E L

CCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCAT 2000
GGTGGTTGGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATGGTAGTA

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L P S

Tuesday, 18 November 1997 10:33 Fig 9 Page 5

fig .Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
CCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGCACC
GGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGG 2100
                                        pHH14-3
                                        pCB210-14
                        full available ORF HU-Unc53/1 = pLM1 OR
S  S  D  T  T  H  A  S  K  V  P  D  L  H  A  T  S  S  A  S  G  G  P  L  P  S  C  F  T  P  S  P  A  P

CATCCTCAATATTAACTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGC
GTAGGAGTTATAATTGAGTCGGTCGAAGAGGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCG 2200
                                        pHH14-3
                                        pCB210-14
                        full available ORF HU-Unc53/1 = pLM1 OR
 I  L  N  I  N  S  A  S  F  S  Q  G  L  E  L  M  S  G  F  S  V  P  K  E  T  R  M  Y  P  K  L  S  G

CTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTC
GACGTGTCCTCGTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAG 2300
                                        pHH14-3
                                        pCB210-14
                        full available ORF HU-Unc53/1 = pLM1 OR
L  H  R  S  M  E  S  L  Q  M  P  M  S  L  P  S  A  F  P  S  S  T  P  V  P  T  P  P  A  P  P  A  A

CCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGG
GGTGTCTTCTTCTCTGCCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCC 2400
                                        pHH14-3
                                                  pHH3b
                              pCB210-14
                        full available ORF HU-Unc53/1 = pLM1 OR
                  rev primer HU53rv3                     rev primer HU53rv2
                                                                  peptide 872628H
P  T  E  E  E  T  E  E  L  T  V  E  G  S  P  R  A  G  Q  L  D  S  N  Q  R  D  R  N  T  L  P  K  K  G
```

F.j 9

Tuesday, 18 November 1997 10:33 Page 6

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
GCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT 2500
CGAGTCCATGGTCGAAGTCAGGGTCCTCCTCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGA
                              pHH14-3
                              pHH3b
rev primer HU53rv1
              full available ORF HU-Unc53/1 = pLM1 OR
 L  R  Y  Q  L  Q  S  Q  E  E  T  K  E  R  R  H  S  H  T  I  G  G  L  P  E  S  D  D  Q  S  E  L  P

TCTCCCCCTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACA 2600
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGT
                              pHH14-3
                              pHH3b
              full available ORF HU-Unc53/1 = pLM1 OR
 S  P  P  A  L  P  M  S  L  S  A  K  G  Q  L  T  N  I  V  S  P  T  A  A  T  T  P  R  I  T  R  S  N

GCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGG 2700
CGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC
                              pHH14-3
                              pHH3b
              full available ORF HU-Unc53/1 = pLM1 OR
 S  I  P  T  H  E  A  A  F  E  L  Y  S  G  S  Q  M  G  S  T  L  S  L  A  E  R  P  K  G  M  I  R  S  G

ATCCTTCCGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCT 2800
TAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGA
              pHH14-3
                              pHH3b
              full available ORF HU-Unc53/1 = pLM1 OR
 S  F  R  D  P  T  D  D  V  H  G  S  V  L  S  L  A  S  S  A  S  S  T  Y  S  S  A  E  E  R  M  Q  S

GAGCAAATCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTT 2900
CTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA
                              U2 ORF = pCB351 ORF
                              pHH3b
              full available ORF HU-Unc53/1 = pLM1 OR
 E  Q  I  R  K  L  R  R  E  L  E  S  S  Q  E  K  V  A  T  L  T  S  Q  L  S  A  N  A  N  L  V  A  A
```

Tuesday, 18 November 1997 10:33 Fi u 9

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence Page 7

TTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGA
AACTCGTCTCGGACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCT 3000

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAAC
GAAAGACTTCTTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCTGTTTTG 3100

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q N

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCTCAACCC 3200

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S V

TCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGA
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCT 3300

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D

CTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTC
GAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGATGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAG 3400

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q H G S T E T A S P S I K S S T L S S V G T D V

Tuesday, 18 November 1997 10:33 Fig 9 Page

Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
ACCG  GCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAG   3500
TGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTC
                               U2 ORF = pCB251 ORF
                                    pHH3b
                      full available ORF HU-Unc53/1 = pLM1 OR
 T  E  G  P  A  H  P  A  P  H  T  R  L  F  H  A  N  E  E  E  E  P  E  K  K  E  V  S  E  L  R  S  E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGT   3600
ATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCA
                               U2 ORF = pCB251 ORF
                                    pHH3b
              peptide B72627H
                      full available ORF HU-Unc53/1 = pLM1 OR
                                U3 ORF = pLM5 ORF
 L  W  E  K  E  M  K  L  T  D  I  R  L  E  A  L  N  S  A  H  Q  L  D  Q  L  R  E  T  M  H  N  M  Q  L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTA   3700
CCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAAT
                               U2 ORF = pCB251 ORF
                                    pHH3b
                      full available ORF HU-Unc53/1 = pLM1 OR
                                U3 ORF = pLM5 ORF
 E  V  D  L  L  K  A  E  N  D  R  L  K  V  A  P  G  P  S  S  G  S  T  P  G  Q  V  P  G  S  S  A  L

TCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTG   3800
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACAC
                               U2 ORF = pCB251 ORF
                                    pHH3b
                      full available ORF HU-Unc53/1 = pLM1 OR
                                U3 ORF = pLM5 ORF
 S  S  P  R  R  S  L  G  L  A  L  T  H  S  F  G  P  S  L  A  D  T  D  L  S  P  M  D  G  I  S  T  C
```

Tuesday, 18 November 1997 10:34 Fig 9

fig 'u-Unc53/1 seq (1 > 6013) Site and Sequence Page 9

```
GTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAG
CAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATC  3900
```

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G C S

```
CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGA
GTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCT  4000
```

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

```
CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATA
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT  4100
```

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V N

```
ACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCT
TGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGA  4200
```

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L

Tuesday, 18 November 1997 10:34 Fig 9 Page 16

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
CCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGC
----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+ 4300
GGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCG
```

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

```
CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACC
----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+ 4400
GCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG
```

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

```
GGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAA
----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+ 4500
CCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTT
```

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C H

Tuesday, 18 November 1997 10:34 Fig 9 Page 11

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCC 4600
CATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72526H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTC 4700
TTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTC 4800
CCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R V L D W V P K L W Y - L H T F L E K H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34 Fig , Page 12

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAG

CACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTC

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T W F I D L V N N S I I P Y L Q E G A K D G I K

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACC 5000

CAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V H G Q K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H

TGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGAT

ACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S D P L H

Tuesday, 18 November 1997 10:34 Fig 9 Page 13

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGTTC 5200
CCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCAAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72825H

A M L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L . G F

GGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAG 5300
CCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTC

pHH3b

pHH15

G N H C H P R T A E R W H Q L S . L L L S P L L F Q S T G S P A P

GAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGG 5400
CTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACC

pHH3b

pHH15

G G E Q E G G G D E R G G T G S V C C T F E N F L G R N G G V A F G

GAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAAC 5500
CTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTG

pHH3b

pHH15

N L C P L N T F T G L L . . L W G K D D S G S F P . L L V S I T N

TCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGG 5600
AGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAGTTTTGTGACGTCGTCAAGGATTTACTAAGAGTGTTCGTTGGGACTCTCTCTGTCAGAACACTCCC

pHH3b

pHH15

W A F W G G V Q H T E . H C S S S . M I . T S N P E R D G L V R

Fig 9.

Tuesday, 18 November 1997 10:34 Page 14
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCCATCACCACTGCCAACAACTCCTCCCCCAGAGATCTGGCTGGAGC 5700
TCTAGACCCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGTCTGGGAAGGGTTAAGGTAGTGGTGACGGTTGTTGAGGAGGGGGTCTCTAGACCGACCTCG
pHH15
E I V G R Q E A P Q I F S Q T L P N S I T T A N N S S P R D L A G A

CCAGAAAAAGAAGCATGTGGTTTAAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAAACAAAAACAAGCAAACAAACAAAAAACAATGGAAA 5800
GGTCTTTTTCTTCGTACACCAAATTTTTTACAAATTTAGTTAGACATTTTCCATTTTTACTTTTTTGTTTTTGTTCGTTTGTTTGTTTTTTGTTACCTTT
Q K K K H V V . K M F K S I C K R . K . K N K N K Q T N K K Q W I

AGATGAAGCTGGAGAGAGAGGAACCAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACC 5900
TCTACTTCGACCTCTCTCTCCTTGGTCAACGGTTCCATCTCTCGACGGGCGAGGACGGGAGACCTACTGTATCCCCTGTAGTTGTTCTGCCGACGGTTGG
R . S V R E R N Q L P R . R A A R S C P L D D I G D I N K T A A N

TGAGAAGTCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAG 6000
ACTCTTCAGTGGTTTGGTGTTTTTATTGGAATGTCGGAAGTCCCTTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTGTTACGTGGCCTTAAGTC
linker?
L R S H Q T T K I T L C P S G K D Y Q L C L S T L . F N N A P E F S

CTTGGACTTAACC 6013
GAACCTGAATTGG
linker?
L D L T
```

11/25/97 02:17PM hh2UNC53_971120 Page 1

```
+1  I  Y  T  D  M  A  N  H  Y  L  T  K  S  G  H  K  R  L  I  K  D  L  Q  Q  D  V
1    ATCTACACAG ACTGGGCCAA TCATTACCTA ACCAAATCCG GCCACAAGCG TCTCATCAAG GATCTCCAGC AAGATGTGAC
     TAGATGTGTC TGACCCGGTT AGTAATGGAT TGGTTTAGGC CGGTGTTCGC AGAGTAGTTC CTAGAGGTCG TTCTACACTG
                                                              97080802
+1  D  G  V  L  L  A  Q  I  I  Q  V  V  A  N  E  K  I  E  D  I  N  G  C  P  K  N
81   AGATGGCGTC CTCCTGGCCC AGATTATCCA GGTTGTGGCA AATGAAAAGA TTGAAGACAT CAATGGCTGT CCGAAGAACA
     TCTACCGCAG GAGGACCGGG TCTAATAGGT CCAACACCGT TTACTTTTCT AACTTCTGTA GTTACCGACA GGCTTCTTGT
               97080801
+1 R  S  Q  M  I  E  N  I  D  A  C  L  N  F  L  A  A  K  G  I  N  I  Q  G  L  S  A
161  GATCCCAAAT GATTGAAAAC ATAGATGCCT GCTTGAATTT CCTGGCAGCT AAGGGAATAA ACATCCAGGG GCTGTCTGCA
     CTAGGGTTTA CTAACTTTTG TATCTACGGA CGAACTTAAA GGACCGTCGA TTCCCTTATT TGTAGGTCCC CGACAGACGT

+1  E  E  I  R  N  G  N  L  K  A  I  L  G  L  F  F  S  L  S  R  Y  K  Q  Q  Q  Q
241  GAAGAGATCA GGAATGGAAA CCTCAAGGCC ATTCTAGGCC TCTTCTTCAG CCTCTCCCGA TACAAGCAGC AGCAGCAGCA
     CTTCTCTAGT CCTTACCTTT GGAGTTCCGG TAAGATCCGG AGAAGAAGTC GGAGAGGGCT ATGTTCGTCG TCGTCGTCGT

+1  P  Q  K  Q  H  L  S  S  P  L  P  P  A  V  S  Q  V  A  G  A  P  S  Q  C  Q  A
321  GCCCCAGAAG CAGCACCTCT CCTCACCTCT GCCGCCCGCC GTATCCCAGG TGGCCGGGGC CCCCTCCCAG TGCCAGGCTG
     CGGGGTCTTC GTCGTGGAGA GGAGTGGAGA CGGCGGGCGG CATAGGGTCC ACCGGCCCCG GGGGAGGGTC ACGGTCCGAC

+1 G  T  P  Q  Q  Q  V  P  V  T  P  Q  A  P  C  Q  P  H  Q  P  A  P  H  Q  Q  S  K
401  GCACCCCTCA GCAGCAGGTG CCAGTCACTC CCCAAGCCCC GTGCCAGCCT CACCAGCCAG CGCCACATCA GCAGTCAAAA
     CGTGGGGAGT CGTCGTCCAC GGTCAGTGAG GGGTTCGGGG CACGGTCGGA GTGGTCGGTC GCGGTGTAGT CGTCAGTTTT

+1  A  Q  A  E  M  Q  S  R  L  S  G  P  T  A  R  V  S  A  A  G  S  E  A  K  T  R
481  GCACAAGCTG AAATGCAGTC CAGACTTTCA GGTCCTACCG CGAGGGTATC CGCTGCAGGC AGCGAGGCCA AAACACGCGG
     CGTGTTCGAC TTTACGTCAG GTCTGAAAGT CCAGGATGGC GCTCCCATAG GCGACGTCCG TCGCTCCGGT TTTGTGCGCC
               97080804
+1  G  S  T  T  A  N  N  R  R  S  Q  S  F  N  N  Y  D  K  S  K  P  V  T  S  P  P
561  AGGGTCAACT ACTGCTAACA ACCGACGCAG CCAGAGCTTT AACAACTATG ATAAATCCAA ACCAGTCACC TCCCCACCCC
     TCCCAGTTGA TGACGATTGT TGGCTGCGTC GGTCTCGAAA TTGTTGATAC TATTTAGGTT TGGTCAGTGG AGGGGTGGGG
          97080803
+1 P  P  P  S  S  H  E  K  E  P  L  A  S  S  A  S  S  H  P  G  M  S  D  N  A  P  A
641  CACCGCCAAG CAGCCACGAG AAAGAGCCTT TGGCAAGTTC AGCCTCCTCC CACCCCGGAA TGAGTGACAA TGCACCTGCT
     GTGGCGGTTC GTCGGTGCTC TTTCTCGGAA ACCGTTCAAG TCGGAGGAGG GTGGGGCCTT ACTCACTGTT ACGTGGACGA

+1  S  L  E  S  G  S  S  S  T  P  T  N  C  S  T  S  S  A  I  P  Q  P  G  A  A  T
721  TCCTTGGAGA GCGGCAGCAG CTCCACCCCT ACTAATTGCA GTACCTCCTC GGCCATCCCG CAGCCCGGTG CAGCCACCAA
     AGGAACCTCT CGCCGTCGTC GAGGTGGGGA TGATTAACGT CATGGAGGAG CCGGTAGGGC GTCGGGCCAC GTCGGTGGTT

+1   P  W  R  S  K  S  L  S  V  K  H  S  A  T  V  S  M  L  S  V  K  P  P  G  P  E
801  GCCTTGGCGC AGCAAATCCC TCAGCGTGAA GCACAGTGCC ACGGTATCCA TGCTCTCGGT CAAGCCTCCT GGGCCTGAGG
     CGGAACCGCG TCGTTTAGGG AGTCGCACTT CGTGTCACGG TGCCATAGGT ACGAGAGCCA GTTCGGAGGA CCCGGACTCC

+1 A  P  R  P  T  P  E  A  M  K  P  A  P  N  N  Q  K  S  M  L  E  K  L  K  L  F  N
881  [illegible] CACACCTGAA GCCATGAAGC CGGCCCCCAA CAATCAGAAG TCCATGCTGG AAAAGCTGAA ACTTTTCAAC
     [illegible] GTGTGGACTT CGGTACTTCG GCCGGGGGTT GTTAGTCTTC AGGTACGACC TTTTCGACTT TGAAAAGTTG

+1  S  K  G  G  S  K  A  G  E  G  P  G  S  R  D  T  S  C  E  R  L  E  T  L  P  S
961  AGTAAAGGGG GCTCAAAGGC AGGTGAGGGG CCGGGGTCCC GGGACACAAG CTGTGAGCGG CTGGAGACTC TGCCCAGCTT
     TCATTTCCCC CGAGTTTCCG TCCACTCCCC GGCCCCAGGG CCCTGTGTTC GACACTCGCC GACCTCTGAG ACGGGTCGAA

+1  E  E  S  E  E  L  E  A  A  S  R  M  L  T  T  V  G  P  A  S  S  S  P  K  I  A
1041 CGAAGAGAGC GAGGAGCTGG AGGCCGCCAG TCGCATGCTC ACCACCGTGG GCCCTGCTTC CAGCAGCCCC AAGATTGCAC
     GCTTCTCTCG CTCCTCGACC TCCGGCGGTC AGCGTACGAG TGGTGGCACC CGGGACGAAG GTCGTCGGGG TTCTAACGTG

+1 L  K  G  I  A  Q  R  T  F  S  R  A  L  T  N  K  K  S  S  L  K  G  N  E  K  E  K
1121 TCAAGGGCAT TGCCCAGAGG ACTTTTAGCC GGGCACTGAC CAACAAGAAG AGTTCTCTGA AAGGCAATGA GAAAGAGAAG
     AGTTCCCGTA ACGGGTCTCC TGAAAATCGG CCCGTGACTG GTTGTTCTTC TCAAGAGACT TTCCGTTACT CTTTCTCTTC

+1  E  K  Q  Q  R  E  K  D  K  E  K  S  K  D  L  A  K  R  A  S  V  T  E  R  L  D
1201 GAGAAACAAC AGCGGGAGAA GGATAAGGAG AAAAGCAAGG ACCTTGCCAA GAGAGCCTCT GTGACGGAGA GGCTGGACCT
     CTCTTTGTTG TCGCCCTCTT CCTATTCCTC TTTTCGTTCC TGGAACGGTT CTCTCGGAGA CACTGCCTCT CCGACCTGGA
```

SEQ ID NO 2: AMINO ACID SEQUENCE OF HU-UNC-53/2 PROTEIN

11/25/97 02:17PM hh2UNC53_971120 Page 2

```
+1   K  E  E  P  K  E  D  P  S  G  A  A  V  P  E  M  P  K  K  S  S  K  I  A  S  F
1281 CAAGGAGGAG CCAAAAGAAG ACCCCAGTGG AGCAGCTGTG CCCGAGATGC CAAAAAAGTC CTCCAAGATT GCCAGCTTCA
     GTTCCTCCTC GGTTTTCTTC TGGGGTCACC TCGTCGACAC GGGCTCTACG GTTTTTTCAG GAGGTTCTAA CGGTCGAAGT

+1 I  P  K  G  G  K  L  N  S  A  K  K  E  P  M  A  P  S  H  S  G  I  P  K  P  G  M
1361 TCCCCAAAGG GGGGAAGCTC AACAGTGCCA AGAAGGAGCC CATGGCCCCT TCCCACAGTG GAATACCAAA ACCAGGAATG
     AGGGGTTTCC CCCCTTCGAG TTGTCACGGT TCTTCCTCGG GTACCGGGGA AGGGTGTCAC CTTATGGTTT TGGTCCTTAC

+1 K  S  M  P  G  K  S  P  S  A  P  A  P  S  K  E  G  E  R  S  R  S  G  K  L  S
1441 AAGAGCATGC CCGGGAAATC CCCAAGTGCC CCAGCGCCTT CCAAGGAAGG GGAGCGGAGC CGGAGTGGGA AGCTGAGCTC
     TTCTCGTACG GGCCCTTTAG GGGTTCACGG GGTCGCGGAA GGTTCCTTCC CCTCGCCTCG GCCTCACCCT TCGACTCGAG

+1   G  L  P  Q  Q  K  P  Q  L  D  G  R  H  S  S  S  S  S  S  L  A  S  S  E  G  K
1521 AGGACTCCCC CAGCAGAAGC CCCAGCTGGA CGGCAGACAC TCCAGTTCCT CTTCCAGCCT GGCGTCCTCA GAAGGAAAAG
     TCCTGAGGGG GTCGTCTTCG GGGTCGACCT GCCGTCTGTG AGGTCAAGGA GAAGGTCGGA CCGCAGGAGT CTTCCTTTTC

+1 G  P  G  G  T  T  L  N  H  S  I  S  S  Q  T  V  S  G  S  V  G  T  T  Q  T  T  G
1601 GCCCAGGAGG GACCACCCTG AACCACAGCA TCAGCAGCCA GACTGTCAGT GGGTCTGTCG GGACCACCCA GACCACAGGA
     CGGGTCCTCC CTGGTGGGAC TTGGTGTCGT AGTCGTCGGT CTGACAGTCA CCCAGACAGC CCTGGTGGGT CTGGTGTCCT

+1 S  N  T  V  S  V  Q  L  P  Q  P  Q  Q  Q  Y  N  H  P  N  T  A  T  V  A  P  F
1681 AGCAATACCG TCAGTGTTCA GCTACCTCAG CCCCAGCAGC AATACAACCA TCCCAACACT GCCACGGTTG CACCTTTCCT
     TCGTTATGGC AGTCACAAGT CGATGGAGTC GGGGTCGTCG TTATGTTGGT AGGGTTGTGA CGGTGCCAAC GTGGAAAGGA

+1   Y  R  S  Q  T  D  T  E  G  N  V  T  A  E  S  S  S  T  G  V  S  V  E  P  S  N
1761 GTACAGGTCT CAGACGGACA CTGAAGGGAA TGTTACTGCC GAGTCAAGCT CAACAGGTGT GAGCGTGGAG CCCAGCCACT
     CATGTCCAGA GTCTGCCTGT GACTTCCCTT ACAATGACGG CTCAGTTCGA GTTGTCCACA CTCGCACCTC GGGTCGGTGA

+1 F  T  K  T  G  Q  P  A  L  E  E  L  T  G  E  D  P  E  A  R  R  L  R  T  V  K  N
1841 TCACCAAGAC TGGACAGCCT GCTCTGGAAG AACTCACTGG GGAAGATCCT GAGGCTCGGC GGCTGCGGAC AGTGAAGAAC
     AGTGGTTCTG ACCTGTCGGA CGAGACCTTC TTGAGTGACC CCTTCTAGGA CTCCGAGCCG CCGACGCCTG TCACTTCTTG

+1   I  A  D  L  R  Q  N  L  E  E  T  M  S  S  L  R  G  T  Q  V  T  H  S  T  L  E
1921 ATCGCTGATC TGCGGCAGAA TTTGGAGGAA ACCATGTCCA GTTTAAGGGG AACTCAGGTT ACACACAGCA CATTGGAAAC
     TAGCGACTAG ACGCCGTCTT AAACCTCCTT TGGTACAGGT CAAATTCCCC TTGAGTCCAA TGTGTGTCGT GTAACCTTTG
                                                  97041702
+1   T  F  D  T  N  V  T  T  E  M  S  G  R  S  I  L  S  L  T  G  R  P  T  P  L  S
2001 CACGTTTGAC ACCAATGTCA CCACGGAGAT GAGTGGCCGT AGCATACTCA GCTTGACAGG GAGGCCCACA CCTCTGTCCT
     GTGCAAACTG TGGTTACAGT GGTGCCTCTA CTCACCGGCA TCGTATGAGT CGAACTGTCC CTCCGGGTGT GGAGACAGGA
                                   97041701
+1 W  R  L  G  Q  S  S  P  R  L  Q  A  G  D  A  P  S  M  G  N  G  Y  P  P  R  A  N
2081 GGAGACTGGG CCAGTCCAGC CCTCGGCTCC AAGCAGGAGA CGCCCCCTCA ATGGGAAATG GGTATCCCCC TCGAGCCAAT
     CCTCTGACCC GGTCAGGTCG GGAGCCGAGG TTCGTCCTCT GCGGGGGAGT TACCCTTTAC CCATAGGGGG AGCTCGGTTA

+1   A  S  R  F  I  N  T  E  S  G  R  Y  V  Y  S  A  P  L  R  R  Q  L  A  S  R  G
2161 GCCAGCAGGT TCATCAACAC TGAGTCAGGT CGCTATGTGT ACTCCGCCCC [illegible] [illegible] [illegible]
     CGGTCGTCCA AGTAGTTGTG ACTCAGTCCA GCGATACACA TGAGGCGGGG [illegible] [illegible] [illegible]

+1   S  V  C  H  V  D  V  S  D  K  A  G  D  E  M  D  L  E  G  I  S  M  D  A  P  G
2241 TAGTGTCTGC CATGTGGACG TCTCAGACAA GGCAGGAGAT GAGATGGACC TGGAAGGCAT CAGCATGGAC GCCCCTGGCT
     ATCACAGACG GTACACCTGC AGAGTCTGTT CCGTCCTCTA CTCTACCTGG ACCTTCCGTA GTCGTACCTG CGGGGACCGA

+1 Y  M  S  D  G  D  V  L  S  K  N  I  R  T  D  D  I  T  S  G  Y  M  T  D  G  G  L
2321 ACATGAGCGA TGGGGATGTT CTGAGCAAGA ACATCCGGAC CGATGACATT ACAAGCGGAT ACATGACTGA TGGTGGACTT
     TGTACTCGCT ACCCCTACAA GACTCGTTCT TGTAGGCCTG GCTACTGTAA TGTTCGCCTA TGTACTGACT ACCACCTGAA

+1 G  L  Y  T  R  R  L  N  R  L  P  D  G  M  A  V  V  R  E  T  L  Q  R  N  T  S
2401 GGCCTCTATA CCCGTCGCCT GAACCGGCTC CCTGATGGGA TGGCTGTGGT ACGGGAGACC CTGCAACGAA ATACCTCCCT
     CCGGAGATAT GGGCAGCGGA CTTGGCCGAG GGACTACCCT ACCGACACCA TGCCCTCTGG GACGTTGCTT TATGGAGGGA

+1   G  L  G  D  A  D  S  W  D  D  S  S  S  V  S  S  G  I  S  D  T  I  D  N  L  S
2481 GGGCCTCGGA GACGCTGACA GCTGGGACGA CAGCAGCTCC GTCAGCAGCG GCATCAGCGA CACCATAGAC AACCTCAGCA
     CCCGGAGCCT CTGCGACTGT CGACCCTGCT GTCGTCGAGG CAGTCGTCGC CGTAGTCGCT GTGGTATCTG TTGGAGTCGT
                           97012801
```

FIGURE 11d

```
11/25/97 01:00PM              hh2UNC53_971120                    Page 3

  +1 T  D  D  I   N  T  S   S  S  I  S   S  Y  A   N  T  P   A  S  S  R   K  N  L   D  V  Q
2561  CTGATGACAT CAACACCAGC TCCTCCATCA GCTCTTATGC CAACACACCT GCCTCCTCTC GAAAAAACCT GGATGTGCAG
      GACTACTGTA GTTGTGGTCG AGGAGGTAGT CGAGAATACG GTTGTGTGGA CGGAGGAGAG CTTTTTTGGA CCTACACGTC
 .......................................................................................
  +1 T  D  A  E   K  H  S   Q  V  E   R  N  S  L   W  S  Q   D  D  V   K  K  S  D   G  Q
2641  ACTGATGCTG AGAAGCACTC ACAGGTGGAG AGGAATTCCC TGTGGTCTGG TGATGATGTC AAGAAATCAG ACGGAGGCTC
      TGACTACGAC TCTTCGTGAG TGTCCACCTC TCCTTAAGGG ACACCAGACC ACTACTACAG TTCTTTAGTC TGCCTCCGAG
 .......................................................................................
  +1  D  S  G   I  K  M  E   P  G  S   K  W  R   R  N  P  S   D  V  S   D  E  S   D  K  S
2721  AGACAGCGGC ATAAAAATGG AGCCAGGTTC CAAGTGGAGG CGGAATCCTT CTGATGTGTC TGACGAGTCC GACAAAAGCA
      TCTGTCGCCG TATTTTTACC TCGGTCCAAG GTTCACCTCC GCCTTAGGAA GACTACACAG ACTGCTCAGG CTGTTTTCGT
 .......................................................................................
  +1 T  S  G  K   K  N  P   V  I  S  Q   T  G  S   W  R  R   G  M  T  A   Q  V  G   I  T  M
2801  CGTCGGGCAA GAAGAATCCT GTCATCTCCC AGACAGGCTC ATGGCGGCGA GGCATGACAG CTCAGGTGGG CATCACCATG
      GCAGCCCGTT CTTCTTAGGA CAGTAGAGGG TCTGTCCGAG TACCGCCGCT CCGTACTGTC GAGTCCACCC GTAGTGGTAC
 .......................................................................................
  +1  P  R  T  K   P  S  A   P  A  G   A  L  K  T   P  G  T   G  K  T   D  D  A  K   V  S
2881  CCAAGGACGA AGCCTTCAGC CCCGGCAGGC GCACTGAAGA CCCCAGGAAC TGGAAAAACA GACGACGCAA AGGTGTCTGA
      GGTTCCTGCT TCGGAAGTCG GGGCCGTCCG CGTGACTTCT GGGGTCCTTG ACCTTTTTGT CTGCTGCGTT TCCACAGACT
 .......................................................................................
  +1  K  G  R   L  S  P  K   A  S  Q   V  K  R   S  P  S  D   A  G  R   S  S  G   D  E  S
2961  GAAAGGAAGG CTTTCTCCTA AAGCCTCCCA GGTGAAGCGC TCCCCATCAG ATGCAGGCCG GAGCAGTGGT GACGAATCCA
      CTTTCCTTCC GAAAGAGGAT TTCGGAGGGT CCACTTCGCG AGGGGTAGTC TACGTCCGGC CTCGTCACCA CTGCTTAGGT
 ...........................................97012802...................................
  +1 K  K  P  L   P  S  S   S  R  T  P   T  A  N   A  N  S   F  G  F  K   K  Q  S   G  S  A
3041  AAAAGCCCCT CCCCAGCAGC TCTAGGACAC CTACTGCCAA TGCCAACAGC TTTGGGTTCA AGAAGCAGAG TGGTTCCGCC
      TTTTCGGGGA GGGGTCGTCG AGATCCTGTG GATGACGGTT ACGGTTGTCG AAACCCAAGT TCTTCGTCTC ACCAAGGCGG
 .......................................................................................
  +1 A  G  L  A   M  I  T   A  S  G   V  T  V  T   S  R  S   A  T  L   G  K  I  P   K  S
3121  GCCGGCCTGG CCATGATCAC AGCCAGCGGG GTGACTGTCA CCAGCAGGTC AGCCACACTG GGCAAAATCC CAAAGTCATC
      CGGCCGGACC GGTACTAGTG TCGGTCGCCC CACTGACAGT GGTCGTCCAG TCGGTGTGAC CCGTTTTAGG GTTTCAGTAG
 .......................................................................................
  +1  A  L  V   S  R  S  A   G  R  K   S  S  H   D  G  A  Q   N  Q  D   D  G  Y   L  A  L
3201  TGCACTCGTC AGTCGGTCTG CTGGTCGGAA GTCAAGTATG GATGGGGCTC AGAATCAGGA TGACGGGTAT CTAGCCCTAA
      ACGTGAGCAG TCAGCCAGAC GACCAGCCTT CAGTTCATAC CTACCCCGAG TCTTAGTCCT ACTGCCCATA GATCGGGATT
 .......................................................................................
  +1 S  S  R  T   N  L  Q   Y  R  S  L   P  R  P   S  K  S   N  S  R  N   G  A  G   N  R  S
3281  GCTCCCGGAC AAACCTTCAG TACCGGAGTT TGCCGAGGCC CAGTAAGTCC AACAGCCGGA ACGGGGCTGG GAACAGGTCT
      CGAGGGCCTG TTTGGAAGTC ATGGCCTCAA ACGGCTCCGG GTCATTCAGG TTGTCGGCCT TGCCCCGACC CTTGTCCAGA
 .......................................................................................
  +1 S  T  S  S   I  D  S   N  I  S   S  K  S  A   G  L  P   V  P  K   L  R  E  P   S  K
3361  AGCACCAGCA GCATAGATTC CAACATTAGC AGCAAGTCCG CAGGCCTGCC AGTGCCCAAA CTGAGGGAGC CTTCCAAAAC
      TCGTGGTCGT CGTATCTAAG GTTGTAATCG TCGTTCAGGC GTCCGGACGG TCACGGGTTT GACTCCCTCG GAAGGTTTTG
 .......................................................................................
  +1  A  L  G   S  S  L  P   G  L  V   N  Q  T   D  K  E  K   G  I  S   S  D  N   E  S  V
3441  AGCCCTAGGC AGCTCTCTAC CAGGTCTGGT CAACCAAACA GATAAGGAGA AAGGCATCTC ATCAGACAAC GAGAGTGTGG
      TCGGGATCCG TCGAGAGATG GTCCAGACCA GTTGGTTTGT CTATTCCTCT TTCCGTAGAG TAGTCTGTTG CTCTCACACC
 .......................................................................................
  +1 A  S  C  N   S  V  K   V  N  P  A   A  Q  P   V  S  S   P  A  Q  T   S  L  Q   P  G  A
3521  CTTCCTGTAA CTCGGTGAAA GTGAATCCGG CAGCCCAGCC TGTGTCCAGT CCGGCTCAGA CCAGTCTCCA GCCTGGAGCC
      GAAGGACATT GAGCCACTTT CACTTAGGCC GTCGGGTCGG ACACAGGTCA GGCCGAGTCT GGTCAGAGGT CGGACCTCGG
 .......................................................................................
  +1 K  Y  P  D   V  A  S   P  T  L   R  R  L  F   G  G  K   P  T  K   Q  V  P  I   A  T
3601  AAGTACCCAG ATGTGGCCTC TCCCACACTC CGCAGACTCT TTGGTGGAAA GCCTACCAAG CAAGTGCCCA TCGCCACAGC
      TTCATGGGTC TACACCGGAG AGGGTGTGAG GCGTCTGAGA AACCACCCTT CGGATGGTTC GTTCACGGGT AGCGGTGTCG
 .......................................................................................
  +1  E  N  M   K  N  S  V   V  I  S   N  P  H   A  T  M  T   Q  Q  G   N  L  D   S  P  S
3681  TGAAAACATG AAAAATTCGG TGGTCATCTC CAATCCTCAT GCCACCATGA CTCAGCAAGG TAACCTAGAC TCCCCGTCAG
      ACTTTTGTAC TTTTTAAGCC ACCAGTAGAG GTTAGGAGTA CGGTGGTACT GAGTCGTTCC ATTGGATCTG AGGGGCAGTC
 .......................................................................................
  +1 G  S  G  V   L  S  S   G  S  S  S   P  L  Y   S  K  N   V  D  L  N   Q  S  P   L  A  S
3761  GCAGTGGCGT CCTGAGCAGT GGGAGCAGCA GTCCTCTCTA CAGCAAGAAT GTGGACCTCA ACCAGTCTCC GCTAGCCTCC
      CGTCACCGCA GGACTCGTCA CCCTCGTCGT CAGGAGAGAT GTCGTTCTTA CACCTGGAGT TGGTCAGAGG CGATCGGAGG
 .......................................................................................
```

```
11/25/97 01:00PM              hh2UNC53_971120                    Page 4

  +1 S  P  S  S   A  H  S   A  P  S   N  S  L  T   W  G  T   N  A  S   S  S  S  A   V  S
3841  AGCCCCAGCT CAGCCCACTC GGCCCCTTCC AACAGCCTCA CCTGGGGCAC CAACGGCAGC AGCTCCTCCG CAGTTAGCAA
      TCGGGGTCGA GTCGGGTGAG CCGGGGAAGG TTGTCGGAGT GGACCCCGTG GTTGCCGTCG TCGAGGAGGC GTCAATCGTT
 .......................................................................................
  +1  D  G  L   G  F  Q  S   V  S  S   L  H  T   S  C  E  S   I  D  I   S  L  S   S  G  G
3921  GGATGGCCTG GGCTTTCAGT CTGTCAGCAG CCTCCACACC AGCTGTGAGT CCATCGACAT CTCCCTCAGC AGTCGACGGG
      CCTACCGGAC CCGAAAGTCA GACAGTCGTC GGAGGTGTGG TCGACACTCA GGTAGCTGTA GAGGGAGTCG TCAGCTGCCC
 .......................................................................................
  +1 V  P  S  H   N  S  S   T  G  L  I   A  S  S   K  D  D   S  L  T  P   F  V  R   T  H  S
4001  TCCCCAGCCA CAATTCTTCC ACTGGCCTCA TCGCCTCCTC CAAGGACGAC TCCTTGACTC CCTTTGTCAG AACTAACAGT
      AGGGGTCGGT GTTAAGAAGG TGACCGGAGT AGCGGAGGAG GTTCCTGCTG AGGAACTGAG GGAAACAGTC TTGATTGTCA
 .......................................................................................
  +1 V  K  T  T   L  S  E   S  P  L   S  S  P  A   A  S  P   K  F  C   K  S  T  L   P  R
4081  GTGAAGACCA CACTGTCAGA AAGCCCTCTC TCTTCCCCTG CTGCTAGCCC TAAGTTCTGC AGAAGTACTC TGCCCAGGAA
      CACTTCTGGT GTGACAGTCT TTCGGGAGAG AGAAGGGGAC GACGATCGGG ATTCAAGACG TCTTCATGAG ACGGGTCCTT
 .......................................................................................
  +1  Q  D  S   D  P  H  L   D  R  N   T  L  P   K  K  G  L   R  Y  T   P  T  S   Q  L  R
4161  ACAGGACAGT GACCCGCACC TTGATAGGAA CACTTTGCCT AAGAAAGGAC TCAGGTATAC TCCTACCTCC CAGCTTCGCA
      TGTCCTGTCA CTGGGCGTGG AACTATCCTT GTGAAACGGA TTCTTTCCTG AGTCCATATG AGGATGGAGG GTCGAAGCGT
 ...............................................97010708...............................
  +1 T  Q  E  D   A  K  E   W  L  R  S   H  S  A   G  G  L   Q  D  T  A   A  H  S   P  F  S
4241  CGCAAGAAGA TGCAAAAGAA TGGTTACGGT CCCATTCTGC AGGAGGCCTT CAGGACACCG CTGCCAATTC CCCCTTTTCC
      GCGTTCTTCT ACGTTTTCTT ACCAATGCCA GGGTAAGACG TCCTCCGGAA GTCCTGTGGC GACGGTTAAG GGGGAAAAGG
 .......97010709........................................................................
  +1 S  G  S  S   V  T  S   P  S  G   T  R  F  N   F  S  Q   L  A  S   P  T  T  V   T  Q
4321  TCTGGCTCCA GCGTGACTTC TCCCTCCGGA ACAAGATTCA ACTTTTCCCA GCTTGCGAGT CCCACCACTG TCACCCAGAT
      AGACCGAGGT CGCACTGAAG AGGGAGGCCT TGTTCTAAGT TGAAAAGGGT CGAACGCTCA GGGTGGTGAC AGTGGGTCTA
 .......................................................................................
  +1  S  L  S   N  P  T  M   L  R  T   H  S  L   S  N  A  D   G  Q  Y   D  P  Y   T  D  S
4401  GAGCTTGTCC AACCCGACCA TGCTGAGGAC TCACAGCCTC TCCAATGCTG ATGGGCAGTA TGATCCATAC ACTGACAGCC
      CTCGAACAGG TTGGGCTGGT ACGACTCCTG AGTGTCGGAG AGGTTACGAC TACCCGTCAT ACTAGGTATG TGACTGTCGG
 .......................................................................................
  +1 R  F  R  N   S  S  H   S  L  D  E   K  S  R   T  M  S   R  S  G  S   F  R  D   G  F  E
4481  GCTTCCGGAA TAGCTCCATG TCCCTGGATG AGAAGAGCAG AACCATGAGC CGTTCAGGCT CATTCCGGGA TGGGTTTGAA
      CGAAGGCCTT ATCGAGGTAC AGGGACCTAC TCTTCTCGTC TTGGTACTCG GCAAGTCCGA GTAAGGCCCT ACCCAAACTT
 .......................................................................................
  +1  E  V  H  G   S  S  L   S  L  V   S  S  T  L   S  V  Y   S  T  P   E  E  K  C   Q  S
4561  GAAGTTCATG GATCCTCACT CTCCCTGGTT TCCAGCACAT TGTCAGTTTA TTCTACACCA GAAGAAAAAT GCCAGTCAGA
      CTTCAAGTAC CTAGGAGTGA GAGGGACCAA AGGTCGTGTA ACAGTCAAAT AAGATGTGGT CTTCTTTTTA CGGTCAGTCT
 .......................................................................................
  +1  I  R  K   L  R  R  E   L  D  A   S  Q  E   K  V  S  A   L  T  T   Q  L  T   A  N  A
4641  GATTCGCAAG CTGCGGCGGG AACTGGATGC CTCCCAGGAG AAAGTTTCAG CTTTGACCAC CCAGCTGACC GCAAATGCTC
      CTAAGCGTTC GACGCCGCCC TTGACCTACG GAGGGTCCTC TTTCAAAGTC GAAACTGGTG GGTCGACTGG CGTTTACGAG
 .......................................................................................
  +1 H  L  V  A   A  F  E   Q  S  L  G   N  M  T   I  R  L   Q  S  L  T   M  T  A   E  Q  K
4721  ACCTTGTGGC TGCCTTTGAA CAGAGTCTTG GTAACATGAC AATCAGGCTC CAGAGTCTGA CCATGACAGC TGAGCAGAAG
      TGGAACACCG ACGGAAACTT GTCTCAGAAC CATTGTACTG TTAGTCCGAG GTCTCAGACT GGTACTGTCG ACTCGTCTTC
 .......................................................................................
  +1  D  S  E  L   N  E  L   R  K  T   I  E  L  L   K  K  Q   N  A  A   A  Q  A  A   I  N
4801  GATTCAGAAC TGAATGAGTT AAGAAAAACC ATTGAGCTGC TAAAGAAACA GAACGCAGCT GCCCAGGCTG CCATTAATGG
      CTAAGTCTTG ACTTACTCAA TTCTTTTTGG TAACTCGACG ATTTCTTTGT CTTGCGTCGA CGGGTCCGAC GGTAATTACC
 .......................................................................................
  +1  V  I  N   T  P  E  L   N  C  K   G  N  G   T  A  Q  S   A  D  L   P  I  P   P  Q  H
4881  AGTAATTAAC ACACCTGAGC TCAACTGCAA AGGAAACGGC ACTGCCCAGT CTGCAGACCT CCCGATCCCG AGGCAGCACT
      TCATTAATTG TGTGGACTCG AGTTGACGTT TCCTTTGCCG TGACGGGTCA GACGTCTGGA GGGCTAGGGC TCCGTCGTGA
 .......................................................................................
  +1 S  S  D  S   V  S  S   I  N  S  A   T  S  H   S  S  V   G  S  N  I   E  S  D   S  K  K
4961  CCTCAGACAG CGTCTCCAGC ATCAACAGTG CCACCAGCCA CTCCAGTGTG GGCAGCAACA TAGAGAGTGA CTCAAAGAAG
      GGAGTCTGTC GCAGAGGTCG TAGTTGTCAC GGTGGTCGGT GAGGTCACAC CCGTCGTTGT ATCTCTCACT GAGTTTCTTC
 .......................................................................................
  +1  K  K  R  K   N  W  L   R  S  S   F  K  Q  A   F  G  K   K  K  S   P  K  S  A   S  S
5041  AAGAAGAGGA AGAACTGGTT ACGCAGCTCC TTCAAGCAAG CTTTCGGGAA GAAGAAGTCC CCAAAATCTG CGTCCTCTTA
      TTCTTCTCCT TCTTGACCAA TGCGTCGAGG AAGTTCGTTC GAAAGCCCTT CTTCTTCAGG GGTTTTAGAC GCAGGAGAAT
 .......................................................................................
```

```
 +1   S  D  I  E  E  T  T  D  S  S  L  P  S  S  P  K  L  P  H  N  G  S  T  G  S  T
5121  TTCAGATATT GAGGAGACGA CGGATTCTTC TTTGCCTTCC TCACCAAAGT TACCCCACAA TGGGTCCACA GGTTCCACCC
      AAGTCTATAA CTCCTCTGCT GCCTAAGAAG AAACGGAAGG AGTGGTTTCA ATGGGGTGTT ACCCAGGTGT CCAAGGTGGG

 +1 P  L  L  R  N  S  H  S  N  S  L  I  S  E  C  H  D  S  F  A  E  T  V  M  Q  L  R
5201  CACTGCTGAG GAATTCTCAC TCCAACTCTC TAATTTCCGA ATGCATGGAT AGTGAAGCTG AGACCGTCAT GCAGCTCCGA
      GTGACGACTC CTTAAGAGTG AGGTTGAGAG ATTAAAGGCT TACGTACCTA TCACTTCGAC TCTGGCAGTA CGTCGAGGCT

 +1   N  E  L  R  D  K  E  M  K  L  T  D  I  R  L  E  A  L  S  S  A  H  Q  L  D  Q
5281  AATGAGTTAA GAGACAAGGA GATGAAGCTG ACGGATATCC GCTTAGAAGC TCTCAGTTCT GCCCACCAGC TGGACCAGCT
      TTACTCAATT CTCTGTTCCT CTACTTCGAC TGCCTATAGG CGAATCTTCG AGAGTCAAGA CGGGTGGTCG ACCTGGTCGA
                                                    EST 46037-R2

 +1    R  E  A  M  N  R  H  Q  S  E  I  E  K  L  K  A  E  N  D  R  L  K  S  E  S  Q
5361  GCGGGAGGCC ATGAACAGGA TGCAGAGTGA AATAGAGAAG CTGAAAGCTG AGAATGATCG GCTGAAGTCA GAGTCTCAAG
      CGCCCTCCGG TACTTGTCCT ACGTCTCACT TTATCTCTTC GACTTTCGAC TCTTACTAGC CGACTTCAGT CTCAGAGTTC
                                                                     EST 46037-R1

 +1 G  S  G  C  S  R  A  P  S  Q  V  S  I  S  A  S  P  R  Q  S  M  G  L  S  Q  H  S
5441  GCAGTGGCTG CAGCCGGGCT CCTTCCCAAG TGTCCATCTC TGCCTCCCCG AGGCAGTCCA TGGGCCTCTC CCAGCACAGC
      CGTCACCGAC GTCGGCCCGA GGAAGGGTTC ACAGGTAGAG ACGGAGGGGC TCCGTCAGGT ACCCGGAGAG GGTCGTGTCG

 +1 L  N  L  T  E  S  T  S  L  D  M  L  L  D  D  T  G  E  C  S  A  R  K  E  G  G
5521  TTGAACCTCA CTGAGTCAAC CAGCCTGGAC ATGTTGCTGG ATGACACTGG TGAATGCTCG GCTCGGAAGG AAGGAGGCAG
      AACTTGGAGT GACTCAGTTG GTCGGACCTG TACAACGACC TACTGTGACC ACTTACGAGC CGAGCCTTCC TTCCTCCGTC

 +1   H  V  K  I  V  V  S  F  Q  E  E  M  K  W  K  E  D  S  R  P  H  L  F  L  I  G
5601  GCATGTTAAG ATAGTTGTCA GCTTTCAGGA GGAAATGAAG TGGAAGGAGG ATTCCAGACC ACACCTCTTT CTTATTGGCT
      CGTACAATTC TATCAACAGT CGAAAGTCCT CCTTTACTTC ACCTTCCTCC TAAGGTCTGG TGTGGAGAAA GAATAACCGA

 +1 C  I  G  V  S  G  K  T  K  W  D  V  L  D  G  V  V  R  R  L  F  K  E  Y  I  I  H
5681  GCATTGGAGT TAGTGGCAAG ACGAAGTGGG ATGTGCTCGA TGGGGTGGTT AGACGGCTGT TCAAAGAATA CATCATTCAT
      CGTAACCTCA ATCACCGTTC TGCTTCACCC TACACGAGCT ACCCCACCAA TCTGCCGACA AGTTTCTTAT GTAGTAAGTA

 +1   V  D  P  V  S  Q  L  G  L  N  S  D  S  V  L  G  Y  S  I  G  E  I  K  R  S  N
5761  GTCGACCCAG TGAGTCAGCT AGGGCTGAAT TCAGACAGCG TTCTTGGCTA CAGCATTGGA GAAATCAAGC GCAGCAACAC
      CAGCTGGGTC ACTCAGTCGA TCCCGACTTA AGTCTGTCGC AAGAACCGAT GTCGTAACCT CTTTAGTTCG CGTCGTTGTG
                                     EST 46037

 +1   S  E  T  P  E  L  L  P  C  G  Y  L  V  G  E  N  T  T  I  S  V  T  V  K  G  L
5841  TTCCGAAACA CCGGAGCTGC TTCCTTGTGG CTATCTGGTT GGAGAGAACA CGACCATCTC AGTGACTGTG AAAGGGCTCG
      AAGGCTTTGT GGCCTCGACG AAGGAACACC GATAGACCAA CCTCTCTTGT GCTGGTAGAG TCACTGACAC TTTCCCGAGC

 +1 A  E  N  S  L  D  S  L  V  F  E  S  L  I  P  K  P  I  L  Q  R  Y  V  S  L  L  I
5921  CAGAAAACAG CCTGGACTCA CTGGTGTTTG AGTCCTTGAT TCCCAAGCCC ATCCTGCAGC GCTACGTCTC CCTCCTGATA
      GTCTTTTGTC GGACCTGAGT GACCACAAAC TCAGGAACTA AGGGTTCGGG TAGGACGTCG CGATGCAGAG GGAGGACTAT

 +1   E  H  R  R  I  I  L  S  G  P  S  G  T  G  K  T  Y  L  A  N  R  L  S  E  Y  I
   1  GAGCACCGTC GGATCATTCT CTCTGGGCCC AGCGGCACTG GGAAAACCTA CCTGGCCAAC CGGCTGTCTG AGTATATAGT
      CTCGTGGCAG CCTAGTAAGA GAGACCCGGG TCGCCGTGAC CCTTTTGGAT GGACCGGTTG GCCGACAGAC TCATATATCA

 +1    L  R  E  G  R  E  L  T  D  G  V  I  A  T  F  N  V  D  H  K  S  S  K  E  L  R
6081  GCTTCGAGAG GGACGGGAGT TGACAGACGG GGTTATCGCC ACCTTTAACG TGGACCATAA GTCCAGCAAG GAATTGCGCC
      CGAAGCTCTC CCTGCCCTCA ACTGTCTGCC CCAATAGCGG TGGAAATTGC ACCTGGTATT CAGGTCGTTC CTTAACGCGG

 +1 Q  Y  L  S  N  L  A  D  Q  C  N  S  E  N  N  A  V  D  M  P  L  V  I  I  L  D  N
6161  AGTACCTGTC CAACCTTGCT GACCAGTGCA ACAGTGAGAA CAATGCTGTG GACATGCCCC TCGTCATCAT CCTGGACAAC
      TCATGGACAG GTTGGAACGA CTGGTCACGT TGTCACTCTT GTTACGACAC CTGTACGGGG AGCAGTAGTA GGACCTGTTG
                                            EST46037.F1

 +1   L  H  H  V  S  S  L  G  E  I  F  N  G  L  L  N  C  K  Y  H  K  C  P  Y  I  I
6241  CTACACCACG TGAGCTCTCT GGGCGAGATC TTCAATGGGC TGCTCAACTG CAAGTACCAC AAATGCCCTT ACATAATTGG
      GATGTGGTGC ACTCGAGAGA CCCGCTCTAG AAGTTACCCG ACGAGTTGAC GTTCATGGTG TTTACGGGAA TGTATTAACC
                                                 EST46037-F2

 +1    T  M  N  Q  A  T  S  S  T  P  N  L  Q  L  H  H  N  F  R  W  V  L  C  A  N  H
6321  CACAATGAAC CAGGCTACCT CTTCGACTCC CAACCTGCAG CTTCACCATA ACTTCAGATG GGTGCTTTGT GCCAACCACA
      GTGTTACTTG GTCCGATGGA GAAGCTGAGG GTTGGACGTC GAAGTGGTAT TGAAGTCTAC CCACGAAACA CGGTTGGTGT
```



```
 +1 T  E  P  V  K  G  F  L  G  R  F  L  R  R  K  L  M  E  T  E  I  S  G  R  V  P  H
6401  CGGAGCCTGT GAAGGGTTTC CTTGGCCGAT TCCTGAGGAG GAAGCTCATG GAAACAGAGA TCAGTGGGCG [illegible]
      GCCTCGGACA CTTCCCAAAG GAACCGGCTA AGGACTCCTC CTTCGAGTAC CTTTGTCTCT AGTCACCCGC [illegible]

 +1   M  E  L  V  K  I  I  D  W  I  P  K  V  W  H  H  L  N  R  F  L  E  A  H  S  S
6481  ATGGAGCTGG TAAAAATCAT TGACTGGATT CCCAAGGTCT [illegible]
      TACCTCGACC ATTTTTAGTA ACTGACCTAA GGGTTCCAGA [illegible]

 +1    D  V  T  I  G  P  R  L  F  L  S  C  P  I  D  V  D  G  S  R  V  W  F  T  D  L
6561  GGACGTCACC ATCGGCCCCC GGCTCTTCCT GTCATGCCCC ATCGATGTGG ACGGCTCGAG AGTGTGGTTC ACCGACTTGT
      CCTGCAGTGG TAGCCGGGGG CCGAGAAGGA CAGTACGGGG TAGCTACACC TGCCGAGCTC TCACACCAAG TGGCTGAACA

 +1 W  N  Y  S  I  I  P  Y  L  L  E  A  V  R  E  G  L  Q  L  Y  G  P  R  A  P  W  E
6641  GGAACTATTC CATTATCCCC TATCTCCTGG AAGCCGTCAG AGAAGGACTC CAGCTCTATG [illegible]
      CCTTGATAAG GTAATAGGGG ATAGAGGACC TTCGGCAGTC TCTTCCTGAG GTCGAGATAC [illegible]

 +1   D  P  A  K  W  V  M  D  T  Y  P  W  A  A  S  P  Q  Q  H  E  W  P  P  L  L  Q
6721  GATCCTGCCA AGTGGGTGAT GGACACATAT CCATGGGCAG CCAGCCCACA ACAGCACGAG TGGCCTCCCC [illegible]
      CTAGGACGGT TCACCCACTA CCTGTGTATA GGTACCCGTC GGTCGGGTGT TGTCGTGCTC ACCGGAGGGG [illegible]

 +1    R  P  E  D  V  G  F  D  G  Y  S  M  P  R  E  G  S  T  S  K  Q  H  P  P  S  P
6801  ACGGCCTGAG GATGTCGGCT TCGACGGCTA CTCCATGCCT CGGGAGGGAT CGACAAGCAA [illegible]
      TGCCGGACTC CTACAGCCGA AGCTGCCGAT GAGGTACGGA GCCCTCCCTA GCTGTTCGTT [illegible]

 +1 A  E  G  D  P  L  M  N  M  L  M  R  L  Q  E  A  A  N  Y  S  S  P  Q  S  Y  D  S
6881  CTGAAGGTGA CCCGCTGATG AACATGCTGA TGAGGCTGCA GGAGGCAGCC AACTACTCCA GCCCCCAGAG [illegible]
      GACTTCCACT GGGCGACTAC TTGTACGACT ACTCCGACGT CCTCCGTCGG TTGATGAGGT CGGGGGTCTC [illegible]
                                      EST 923793

 +1   D  S  N  S  N  S  H  H  E  D  I  L  D  S  S  L  E  S  T  L  *  Q  G  P  S  A
6961  GACTCCAACA GCAACAGCCA TCACGAAGAC ATCTTGGACT CCTCTTTGGA GTCCACTCTG TGACAGGGGC [illegible]
      CTGAGGTTGT CGTTGTCGGT AGTGCTTCTG TAGAACCTGA GGAGAAACCT CAGGTGAGAC ACTGTCCCCG [illegible]

 +1    R  P  P  L  L  L  T  A  F  H  L  H  P  P  H  H  P  E  D  D  F  L  S  Q  F  F
7041  GCGCCCTCCT CTTCTCCTCA CCGCATTCCA CCTGCATCCC CCACATCACC CTGAAGATGA CTTCCTGAGC [illegible]
      CGCGGGAGGA GAAGAGGAGT GGCGTAAGGT GGACGTAGGG GGTGTAGTGG GACTTCTACT GAAGGACTCG [illegible]

 +1 A  T  A  L  E  L  R  E  H  R  D  P  P  S  F  S  L  D  L  G  A  G  I  P  G  Q  L
7121  CCACAGCCTT AGAGCTGCGG GAACACCGAG ACCCCCCGTC CTTCAGCCTC GACCTGGGTG CAGGCATCCC [illegible]
      GGTGTCGGAA TCTCGACGCC CTTGTGGCTC TGGGGGGCAG GAAGTCGGAG CTGGACCCAC GTCCGTAGGG [illegible]

 +1   P  A  D  R  F  L  P  Q  R  E  L  H  Y  L  L  L  Y  F  N  Y  C  F  A  L  L  L
7201  CCTGCGGACC GCTTCCTTCC ACAGCGAGAA CTGCACTACC [illegible]
      GGACGCCTGG CGAAGGAAGG TGTCGCTCTT GACGTGATGG [illegible]

 +1    P  P  *  D  T  E  D  T  S  R  E  R  I  I  A  V  I  M  K  K  K  K  K  K  K  K
7281  [illegible] AAAGAAAAA AAAAAAAAAA AAAAAAAAAA
      [illegible] TTTACTTTTT TTTTTTTTTT TTTTTTTTTT

 +1 K  K  K  K
7361  AAAAAAAAAA AAA
      TTTTTTTTTT TTT
```

SEQ. ID No 3 — Nucleic Acid Sequence of hu-unc-53/1 gene

Page 1

GATATCTGCAGAATTCGGCTTCTTTGAGCAAGTTCAGCCTGGTTAAGTCCAAGCTGAATTCCGGGGAAAGCCGAGCCGGATCCCTCGAGGACCCTATGCG 100
CTATAGACGTCTTAAGCCGAAGAAACTCGTTCAAGTCGGACCAATTCAGGTTCGACTTAAGGCCCCTTTCGGCTCGGCCTAGGGAGCTCCTGGGATACGC

pCR2.1 linker — lambda gt10 primer EcoRI — suspect sequence linker? — pHH14-3

I S A E F G F F E Q V Q P G . V Q A E F R G K P S R I P R G P Y A

GAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGCGGCCGTGAGCGAAGATGGCAAATCGGACGACGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCT 200
CTCCAGTTCGGCGAGTCGTTCCGCGGACTTCGCCGGCACTCGCTTCTACCGTTTAGCCTGCTGCTCGACGAGAGGTCGTTCCGGTTCCGCGTTTTCTCGA

pHH14-3

E V K P L S K A P E A A V S E D G K S D D E L L S S K A K A Q K S

CTGGGCCTGTCCCCTCTGCCAAGGGCCAGGAGGAGCGCGCCTTCCTCAAGGTGGACCCCGAGCTGGTGGTGACCGTGCTGGGAGACCTGGAGCAGCTGCT 300
GACCCGGACAGGGGAGACGGTTCCCGGTCCTCCTCGCGCGGAAGGAGTTCCACCTGGGGCTCGACCACCACTGGCACGACCCTCTGGACCTCGTCGACGA

pHH14-3

S G P V P S A K G Q E E R A F L K V D P E L V V T V L G D L E Q L L

CTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGA 400
GAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCGGACGCT

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S L R

GGGTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTC 500
CCCAGGGTCCACTGAGTGTCGAGGGACCTCTACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAG

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

G S Q V T H S S L E M T C Y D S D D A N P R S V S S L S N R S S P

TGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTACAT 600
ACAGTACCGCGATACCGGTCAGGTCAGGCGCCGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGACCATGTA

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

L S W R Y G Q S S P R L Q A G D A P S V G G S C R S E G T P A W Y [illegible]

Tuesday, 18 November 1997 10:33 Fig 9 Page 2
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
GCACGGCGAACGGGCCCACTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCG 700
CGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGC
                              pHH14-3
                   ORF (1-579bp) = pLM7 ORF
            full available ORF HU-Unc53/1 = pLM1 OR
 H  G  E  R  A  H  Y  S  H  T  M  P  M  R  S  P  S  K  L  S  H  I  S  R  L  E  L  V  E  S  L  D  S

GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCA 800
CTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTACTTTCGT
                              pHH14-3
                   ORF (1-579bp) = pLM7 ORF
            full available ORF HU-Unc53/1 = pLM1 OR
 D  E  V  D  L  K  S  G  Y  M  S  D  S  D  L  M  G  K  T  M  T  E  D  D  D  I  T  T  G  W  D  E  S

GCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC 900
CGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACG
                              pHH14-3
                                                      pCB212
                   ORF (1-579bp) = pLM7 ORF
            full available ORF HU-Unc53/1 = pLM1 OR
 S  S  I  S  S  G  L  S  D  A  S  D  N  L  S  S  E  E  F  N  A  S  S  S  L  N  S  L  P  S  T  P  T  A

TTCTCGCAGGAACTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCC 1000
AAGAGCGTCCTTGAGTTGTTATCACGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTCTTTCGG
                              pHH14-3
                              pCB212
            full available ORF HU-Unc53/1 = pLM1 OR
 S  R  R  N  S  T  I  V  L  R  T  D  S  E  K  R  S  L  A  E  S  G  L  S  W  F  S  E  S  E  E  K  A

CCTAAAAAACTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCATCCAAGG 1100
GGATTTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCC
                              pHH14-3
                              pCB212
            full available ORF HU-Unc53/1 = pLM1 OR
 P  K  K  L  E  Y  D  S  G  S  L  K  M  E  P  G  T  S  K  W  R  R  E  R  P  E  S  C  D  D  S  S  K
```

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site ar sequence

Fig 9

Page 3

GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCACACAGC
CACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCG 1200

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGAT
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGGAGACTA 1300

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCCTGAGTGATGCTAAGAAGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCA
CGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT 1400

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC
GTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCTG 1500

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R K T

TAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGT
ATCGAATCTACAAAGGTTGTCACGTCTCGGTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCA 1600

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A K S S

Tuesday, 18 November 1997 10:33

Fig 9

Page 7

fig Hu-Unc53/1 seq (1 > 6013) Site a. Sequence

TCTATGAGCGTGACCGGCGGGCGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCA 1700
AGATACTCGCACTGGCCGCCCGCCCCACCTGGAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGGT

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HU53rv4

S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T P S

GACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGC 1800
CTGACTTCCTCGGATGGTTCCATCGGTCACCCGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCGTCACCG

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

R L K E P T K V A S G R T T P A P V N Q T D R E K E K A K A K A V A

CTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCAGAGCTG 1900
GAACCTGAGTCTGTTGTAGAGGAACTTCTCATAACCGAGGGGTCTTTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGTCTCGAC

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A E L

CCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCAT 2000
GGTGGTTGGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCAGTTGAGGTTGTCAGACCTAGATGGTAGTA

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L P S

Tuesday, 18 November 1997 10:33 Fig 9 Page 5
fig Hu-Unc53/1 seq (1 > 6013) Site , Sequence

```
CCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGCACC 2100
GGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGG
                                   pHH14-3
                                   pCB210-14
                    full available ORF HU-Unc53/1 = pLM1 OR
 S  S  D  T  T  H  A  S  K  V  P  D  L  H  A  T  S  S  A  S  G  G  P  L  P  S  C  F  T  P  S  P  A  P

CATCCTCAATATTAACTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGC 2200
GTAGGAGTTATAATTGAGTCGGTCGAAGAGGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCG
                                   pHH14-3
                                   pCB210-14
                    full available ORF HU-Unc53/1 = pLM1 OR
 I  L  N  I  N  S  A  S  F  S  Q  G  L  E  L  M  S  G  F  S  V  P  K  E  T  R  M  Y  P  K  L  S  G

CTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTC 2300
GACGTGTCCTCGTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAG
                                   pHH14-3
                                   pCB210-14
                    full available ORF HU-Unc53/1 = pLM1 OR
 L  H  R  S  M  E  S  L  Q  M  P  M  S  L  P  S  A  F  P  S  S  T  P  V  P  T  P  P  A  P  P  A  A

CCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGG 2400
GGTGTCTTCTTCTCTGCCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCC
                                   pHH14-3
                                         pHH3b
                             pCB210-14
                    full available ORF HU-Unc53/1 = pLM1 OR
             rev primer HU53rv3                    rev primer HU53rv2
                                                          peptide S72628H
 P  T  E  E  E  T  E  E  L  T  V  E  G  S  P  R  A  G  Q  L  D  S  N  C  R  D  R  N  T  L  P  K  N  G
```

Fig 9

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and sequence

Page 6

GCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT
CGAGTCCATGGTCGAAGTCAGGGTCCTCCTCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGA 2500

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L P

TCTCCCCCTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACA
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGT 2600

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S N

GCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGG
CGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC 2700

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R S G

ATCCTTCCGAGACCCCACGGACGATGTTCACGTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCT
TAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGA 2800

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

I F R D P T D D V H V S V L S L A S S A S S T Y S S A E E R M Q S

GAGCAAATCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCT
CTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGA 2900

U2 ORF = pCB351 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

E Q I R K L R R E L E S S Q E K V A T L T S Q L S A N A N L V A A

Tuesday, 18 November 1997 10:33 Fig 9 Page

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

TTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGA 3000
AACTCGTCTCGGACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCT

U2 ORF = pCS251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAAC 3100
GAAAGACTTCTTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCTGTTTTG

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q N

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGG 3200
AGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCTCAACCC

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S W

TCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGA 3300
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCT

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D

CTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTC 3400
GAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGATGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAG

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q - G S T E T A S P S I K S S T L S S V G T D V

Tuesday, 18 November 1997 10:33

Fig 9

Page 8

Hu-Unc53/1 seq (1 > 6013) Site and Sequence

ACC[illegible]GCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGC
TGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTCG 3500

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTT
ATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAA 3600

U2 ORF = pCB251 ORF

pHH3b

peptide B72627H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTA
CCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAAT 3700

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTG
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACAC 3800

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S T C

Tuesday, 18 November 1997 10:34 Fig 9 Page 9

fig 'u-Unc53/1 seq (1 > 6013) Site and Sequence

GTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAG 3900
CAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATC

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGA 4000
GTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCT

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATA 4100
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V N

ACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGGGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCT 4200
TGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACCCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N I S V S L K G L K E K W V D S L V F E T L I P K P M M Q H Y I S L

Fig. 9

Tuesday, 18 November 1997 10:34
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence
Page 10

CCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGC 4300
GGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CCTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACC 4400
GGACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

GGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAA 4500
CCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C K

Tuesday, 18 November 1997 10:34 Fig 9 Page 11

fig Hu-Unc53/1 seq (1 > 6013) Site Sequence

GTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCC 4600
CATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72629H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTC 4700
TTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGGTGCTCGACTGGGTACCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTC 4800
CCCACGAGCTGACCCATGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

F V L S W V F K L W Y - _ H T F L E K H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34

fig . Hu-Unc53/1 seq (1 > 6013) Site an. equence

Fig. 9

age 12

GTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAC
CACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTG 4900

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G I I

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACC
CAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGG 5000

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V - G C K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H

TGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGAT
ACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTA 5100

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

- P P P T V G P - S I A E P P E D R T V K D S T P S S L D S C P L

Tuesday, 18 November 1997 10:34 *Fig. 9* Page 13

fig Hu-Unc53/1 seq (1 > 6013) Site a Sequence

GGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTC
CCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAG 5200

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72625H

A M L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L . G F

GGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAG
CCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTC 5300

pHH3b

pHH15

G N H C H P R T A E R W H Q L S . L L L S P L L F Q S T G S P A P

GAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGG
CTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACC 5400

pHH3b

pHH15

G G E Q E G G G D E R G G T G S W C C T F E N F L G R N G G V A F G

GAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAAC
CTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTG 5500

pHH3b

pHH15

N L C P L N T F T G L L . . L V G K D D S G S F F . L L V S I T N

TCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATTAAAACACTGCAGCAGTTCCTAAATGATTCTCAGAAGCAACCCTGAGAGAGACAGTCTTGTGAGGG
AGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAATTTTGTGACGTCGTCAAGGATTTACTAAGAGTGTTCGTTGGGACTCTCTCTGTCAGAACACTCCC 5600

pHH3b

pHH15

[illegible]

Tuesday, 18 November 1997 10:34 Fig 9. Page 1†

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCCATCACCACTGCCAACAACTCCTCCCCCAGAGATCTGGCTGGAGC 5700
TCTAGACCCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGTCTGGGAAGGGTTAAGGTAGTGGTGACGGTTGTTGAGGAGGGGGTCTCTAGACCGACCTCG

pHH15

E I W G R Q E A P Q I F S Q T L P N S I T T A N N S S P R D L A G A

CCAGAAAAAGAAGCATGTGGTTTAAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAAACAAAAACAAGCAAACAAACAAAAAACAATGGAAA 5800
GGTCTTTTTCTTCGTACACCAAATTTTTTACAAATTTAGTTAGACATTTTCCATTTTTACTTTTTTGTTTTTGTTCGTTTGTTTGTTTTTTGTTACCTTT

Q K K K H V V . K M F K S I C K R . K . K N K N K Q T N K K Q W I

AGATGAAGCTGGAGAGAGAGGAACCAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACC 5900
TCTACTTCGACCTCTCTCTCCTTGGTCAACGGTTCCATCTCTCGACGGGCGAGGACGGGAGACCTACTGTATCCCCTGTAGTTGTTCTGCCGACGGTTGG

R . S V R E R N Q L P R . R A A R S C P L D D I G D I N K T A A N

TGAGAAGTCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAG 6000
ACTCTTCAGTGGTTTGGTGTTTTTATTGGAATGTCGGAAGTCCCTTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTGTTACGTGGCCTTAAGTC

linker?

L R S H Q T T K I T L Q P S G K D Y Q L C L S T L . F N N A P E F S

CTTGGACTTAACC 6013
GAACCTGAATTGG

linker?

L D L T

FIGURE 11d

```
11/25/97 02:17PM            hh2UNC53_971120                  Page 1

 +1    I  T  I  L  W  A  H  H  Y  L  T  K  S  G  H  K  R  L  I  K  D  L  Q  Q  D  V
    1  [illegible] ACTGGGCCAA TCATTACCTA ACCAAATCCG GCCACAAGCG TCTCATCAAG GATCTCCAGC AAGATGTGAC
       [illegible] TGACCCGGTT AGTAATGGAT TGGTTTAGGC CGGTGTTCGC AGAGTAGTTC CTAGAGGTCG TTCTACACTG
                                                                    97080802
 +1    D  G  V  L  L  A  Q  I  I  Q  V  V  A  H  E  K  I  E  D  I  H  G  C  P  K  N
   81  [illegible] CTCCTGGCCC AGATTATCCA [illegible] AATGAAAAGA TTGAAGACAT CAATGGCTGT CCGAAGAACA
       [illegible] GAGGACCGGG TCTAATAGGT CCAAC[illegible]CGT TTACTTTTCT AACTTCTGTA GTTACCGACA GGCTTCTTGT
              97080801
 +1    [illegible] Q  H  I  E  D  I  D  A  C  L  H  F  L  A  A  K  G  I  H  I  Q  G  L  S  A
  161  [illegible] GATTGAAAAC ATAGATGCCT GCTTGAATTT CCTGGCAGCT AAGGGAATAA ACATCCAGGG GCTGTCTGCA
       CTAGGTTTTA CTAACTTTTG TATCTACGGA CGAACTTAAA GGACCGTCGA TTCCCTTATT TGTAGGTCCC CGACAGACGT

 +1    E  E  I  R  N  G  N  L  K  A  I  L  G  L  F  F  S  L  S  R  Y  K  Q  Q  Q  Q
  241  [illegible] GGAATGGAAA CCTCAAGGCC ATTCTAGGC[illegible] TCTTCTTCAG CCTCTCCCGA TACAAGCAGC AGCAGCAGCA
       [illegible] CCTTACCTTT GGAGTTCCGG TAAGATCCG[illegible] AGAAGAAGTC GGAGAGGGCT ATGTTCGTCG TCGTCGTCGT

 +1    P  Q  K  Q  H  L  S  S  P  L  P  P  A  V  S  Q  V  A  G  A  P  S  Q  C  Q  A
  321  [illegible] CAGCACCTGT [illegible] [illegible] GTATCCCAGG TGGCCGGGGC CCCCTCCCAG TGCCAGGCTG
       [illegible] [illegible] [illegible] [illegible] CATAGGGTCC ACCGGCCCCG GGGGAGGGTC ACGGTCCGAC

 +1    [illegible] I  P  Q  Q  Q  V  P  V  P  P  Q  A  P  P  Q  P  H  Q  P  A  P  H  Q  Q  S  K
  401  [illegible] [illegible] CCAGCCACTC CCCAAGCCCC GCCTCAGCCT CACCAGCCAG CGCCACATCA GCAGTCAAAA
       [illegible] [illegible] GGTCGGTGAG GGGTTCGGGG CGGAGTCGGA GTGGTCGGTC GCGGTGTAGT CGTCAGTTTT

 +1    A  Q  A  E  M  Q  S  R  L  S  G  P  T  A  R  V  S  A  A  G  S  E  A  K  T  R
  481  [illegible] AAATGCAGTC CAGACTTTCA GGTCCTACCG CGAGGGTATC CGCTGCAGGC AGCGAGGCCA AAACACGCGG
       [illegible]CAC TTTACGTCAG GTCTGAAAGT CCA[illegible] GCTCCCATAG GCGACGTCCG TCGCTCCGGT TTTGTGCGCC
              97080804
 +1    G  S  T  T  A  H  H  K  K  S  Q  S  F  H  H  Y  D  K  S  K  P  V  T  S  P  P
  561  AGGGTCAACT ACTGCTAACA ACCGACGCAG CCAGAGCTTT AACAACTATG ATAAATCCAA ACCAGTCACC TCCCCACCCC
       TCCCAGTTGA TGACGATTGT TGG[illegible]CGTC GGTCTCGAAA TTGTTGATAC TATTTAGGTT TGGTCAGTGG AGGGGTGGGG
         97080803
 +1    P  P  P  S  S  H  E  K  E  P  L  A  S  S  A  S  S  H  P  G  M  S  D  N  A  P  A
  641  CACCGACAAG CAGCCACGAG AAAGAGCCTT TGGCAAGTTC AGCCTCCTCC CACCCCGGAA TGAGTGACAA TGCACCTGCT
       GTGGCTGTTC GTCGGTGCTC TTTCTCGGAA ACCGTTCAAG TCGGAGGAGG GTGGGGCCTT ACTCACTGTT ACGTGGACGA

 +1    S  L  E  S  G  S  S  S  T  P  T  N  C  S  T  S  S  A  I  P  Q  P  G  A  A  T
  721  TCCTTGGAGA GCGGCAGCAG CTCCACCCCT ACTAATTGCA GTACCTCCTC GGCCATCCCG CAGCCCGGTG CAGCCACCAA
       AGGAACCTCT CGCCGTCGTC GAGGTGGGGA TGATTAACGT CATGGAGGAG CCGGTAGGGC GTCGGGCCAC GTCGGTGGTT

 +1    P  W  R  S  K  S  L  S  V  K  H  S  A  T  V  S  M  L  S  V  K  P  P  G  P  E
  801  GCCTTGGAGG AGCAAATCCC TCAGCGTGAA GCACAGTGCC ACGGTATCCA TGCTCTCGGT CAAGCCTCCT GGGCCTGAGG
       CGGAACCTCC TCGTTTAGGG AGTCGCACTT CGTGTCACGG TGCCATAGGT ACGAGAGCCA GTTCGGAGGA CCCGGACTCC

 +1    A  P  R  P  T  P  E  A  M  K  P  A  P  N  N  Q  K  S  M  L  E  K  L  K  L  F  N
  881  CCCCCAGGCC CACACCTGAA GCCATGAAGC CGGCCCCCAA CAATCAGAAG TCCATGCTGG AAAAGCTGAA ACTTTTCAAC
       GGGGGTCCGG GTGTGGACTT CGGTACTTCG GCCGGGGGTT GTTAGTCTTC AGGTACGACC TTTTCGACTT TGAAAAGTTG

 +1    S  K  G  G  S  K  A  G  E  G  P  G  S  R  D  T  S  C  E  R  L  E  T  L  P  S
  961  AGTAAAGGGG GCTCAAAGGC AGGTGAGGGC CCGGGGTCCC GGGACACAAG CTGTGAGCGG CTGGAGACTC TGCCCAGCTT
       TCATTTCCCC CGAGTTTCCG TCCACTCCCG GGCCCCAGGG CCCTGTGTTC GACACTCGCC GACCTCTGAG ACGGGTCGAA

 +1    E  E  S  E  E  L  E  A  A  S  R  M  L  T  T  V  G  P  A  S  S  S  P  K  I  A
 1041  CGAAGAGAGC GAGGAGCTGG AGGCCGCCAG TCGCATGCTC ACCACCGTGG GCCCTGCTTC CAGCAGCCCC AAGATTGCAC
       GCTTCTCTCG CTCCTCGACC TCCGGCGGTC AGCGTACGAG TGGTGGCACC CGGGACGAAG GTCGTCGGGG TTCTAACGTG

 +1    L  K  G  I  A  Q  R  T  F  S  R  A  L  T  H  K  K  S  S  L  K  G  N  E  K  E  K
 1121  TCAAGGGCAT TGCCCAGAGG ACTTTTAGCC GGGCACTGAC CAACAAGAAG AGTTCTCTGA AAGGCAATGA GAAAGAGAAG
       AGTTCCCGTA ACGGGTCTCC TGAAAATCGG CCCGTGACTG GTTGTTCTTC TCAAGAGACT TTCCGTTACT CTTTCTCTTC

 +1    E  K  Q  Q  R  E  K  D  K  E  K  S  K  D  L  A  K  R  A  S  V  T  E  R  L  D
 1201  GAGAAACAAC AGCGGGAGAA GGATAAGGAG AAAAGCAAGG ACCTTGCCAA GAGAGCCTCT GTGACGGAGA GGCTGGACCT
       CTCTTTGTTG TCGCCCTCTT CCTATTCCTC TTTTCGTTCC TGGAACGGTT CTCTCGGAGA CACTGCCTCT CCGACCTGGA
```

SEQ. ID. No 4 Nucleic Acid SEQUENCE of human - unc-53/2 gene

```
11/25/97 02:17PM            hh2UNC53_971120                  Page 2

 +1    K  E  E  P  K  E  D  P  S  G  A  A  V  P  E  M  P  K  K  S  S  K  I  A  S  F
 1281  CAAGGAGGAG CCAAAAGAAG ACCCCAGTGG AGCAGCTGTG CCCGAGATGC CAAAAAAGTC CTCCAAGATT GCCAGCTTCA
       GTTCCTCCTC GGTTTTCTTC TGGGGTCACC TCGTCGACAC GGGCTCTACG GTTTTTTCAG GAGGTTCTAA CGGTCGAAGT

 +1    I  P  K  G  G  K  L  N  S  A  K  K  E  P  M  A  P  S  H  S  G  I  P  K  P  G  M
 1361  TCCCCAAAGG GGGGAAGCTC AACAGTGCCA AGAAGGAGCC CATGGCCCCT TCCCACAGTG GAATACCAAA ACCAGGAATG
       AGGGGTTTCC CCCCTTCGAG TTGTCACGGT TCTTCCTCGG GTACCGGGGA AGGGTGTCAC CTTATGGTTT TGGTCCTTAC

 +1    K  S  M  P  G  K  S  P  S  A  P  A  P  S  K  E  G  E  R  S  R  S  G  K  L  S
 1441  AAGAGCATGC CCGGGAAATC CCCAAGTGCC CCAGCGCCTT CCAAGGAAGG GGAGCGGAGC CGGAGTGGGA AGCTGAGCTC
       TTCTCGTACG GGCCCTTTAG GGGTTCACGG GGTCGCGGAA GGTTCCTTCC CCTCGCCTCG GCCTCACCCT TCGACTCGAG

 +1    G  L  P  Q  Q  K  P  Q  L  D  G  R  H  S  S  S  S  S  S  L  A  S  S  E  G  K
 1521  AGGACTCCCC CAGCAGAAGC CCCAGCTGGA CGGCAGACAC TCCAGTTCCT CTTCCAGCCT GGCGTCCTCA GAAGGAAAAG
       TCCTGAGGGG GTCGTCTTCG GGGTCGACCT GCCGTCTGTG AGGTCAAGGA GAAGGTCGGA CCGCAGGAGT CTTCCTTTTC

 +1    G  P  G  G  T  T  L  N  H  S  I  S  S  Q  T  V  S  G  S  V  G  T  T  Q  T  T  G
 1601  GCCCAGGAGG GACCACCCTG AACCACAGCA TCAGCAGCCA GACTGTCAGT GGGTCTGTCG GGACCACCCA GACCACAGGA
       CGGGTCCTCC CTGGTGGGAC TTGGTGTCGT AGTCGTCGGT CTGACAGTCA CCCAGACAGC CCTGGTGGGT CTGGTGTCCT

 +1    S  N  T  V  S  V  Q  L  P  Q  P  Q  Q  Q  Y  N  H  P  N  T  A  T  V  A  P  F
 1681  AGCAATACCG TCAGTGTTCA GCTACCTCAG CCCCAGCAGC AATACAACCA TCCCAACACT GCCACGGTTG CACCTTTCCT
       TCGTTATGGC AGTCACAAGT CGATGGAGTC GGGGTCGTCG TTATGTTGGT AGGGTTGTGA CGGTGCCAAC GTGGAAAGGA

 +1    Y  R  S  Q  T  D  T  E  G  N  V  T  A  E  S  S  S  T  G  V  S  V  E  P  S  H
 1761  GTACAGGTCT CAGACGGACA CTGAAGGGAA TGTTACTGCC GAGTCAAGCT CAACAGGTGT GAGCGTGGAG CCCAGCCACT
       CATGTCCAGA GTCTGCCTGT GACTTCCCTT ACAATGACGG CTCAGTTCGA GTTGTCCACA CTCGCACCTC GGGTCGGTGA

 +1    F  T  K  T  G  Q  P  A  L  E  E  L  T  G  E  D  P  E  A  R  R  L  R  T  V  K  N
 1841  TCACCAAGAC TGGACAGCCT GCTCTGGAAG AACTCACTGG GGAAGATCCT GAGGCTCGGC GGCTGCGGAC AGTGAAGAAC
       AGTGGTTCTG ACCTGTCGGA CGAGACCTTC TTGAGTGACC CCTTCTAGGA CTCCGAGCCG CCGACGCCTG TCACTTCTTG

 +1    I  A  D  L  R  Q  N  L  E  E  T  M  S  S  L  R  G  T  Q  V  T  H  S  T  L  E
 1921  ATCGCTGATC TGCGGCAGAA TTTGGAGGAA ACCATGTCCA GTTTAAGGGG AACTCAGGTT ACACACAGCA CATTGGAAAC
       TAGCGACTAG ACGCCGTCTT AAACCTCCTT TGGTACAGGT CAAATTCCCC TTGAGTCCAA TGTGTGTCGT GTAACCTTTG
                                              97041702
 +1    T  F  D  T  N  V  T  T  E  M  S  G  R  S  I  L  S  L  T  G  R  P  T  P  L  S
 2001  CACGTTTGAC ACCAATGTCA CCACGGAGAT GAGTGGCCGT AGCATACTCA GCTTGACAGG GAGGCCCACA CCTCTGTCCT
       GTGCAAACTG TGGTTACAGT GGTGCCTCTA CTCACCGGCA TCGTATGAGT CGAACTGTCC CTCCGGGTGT GGAGACAGGA
                                97041701
 +1    W  R  L  G  Q  S  S  P  R  L  Q  A  G  D  A  P  S  M  G  N  G  Y  P  P  R  A  N
 2081  GGAGACTGGG CCAGTCCAGC CCTCGGCTCC AAGCAGGAGA CGCCCCCTCA ATGGGCAATG GGTATCCCCC TCGAGCCAAC
       CCTCTGACCC GGTCAGGTCG GGAGCCGAGG TTCGTCCTCT GCGGGGGAGT TACCCGTTAC CCATAGGGGG AGCTCGGTTG

 +1    A  S  R  P  I  N  T  E  S  G  R  Y  V  Y  S  A  P  L  R  R  Q  L  A  S  R  G
 2161  GCCAGCAGGT TCATCAACAC TGAGTCAGGT CGCTATGTGT ACTCCGCCCC TCTGAGAAGG CAGCTGGCCT CCCGGGGCAG
       CGGTCGTCCA AGTAGTTGTG ACTCAGTCCA GCGATACACA TGAGGCGGGG AGACTCTTCC GTCGACCGGA GGGCCCCGTC

 +1    S  V  C  H  V  D  V  S  D  K  A  G  D  E  M  D  L  E  G  I  S  M  D  A  P  G
 2241  TAGTGTCTGC CATGTGGACG TCTCAGACAA GGCAGGAGAT GAGATGGACC TGGAAGGCAT CAGCATGGAC GCCCCCGGCT
       ATCACAGACG GTACACCTGC AGAGTCTGTT CCGTCCTCTA CTCTACCTGG ACCTTCCGTA GTCGTACCTG CGGGGGCCGA

 +1    Y  M  S  D  G  D  V  L  S  K  N  I  R  T  D  D  I  T  S  G  Y  M  T  D  G  G  L
 2321  ACATGAGCGA TGGGGATGTT CTGAGCAAGA ACATCCGGAC CGATGACATT ACAAGCGGAT ACATGACTGA TGGTGGACTT
       TGTACTCGCT ACCCCTACAA GACTCGTTCT TGTAGGCCTG GCTACTGTAA TGTTCGCCTA TGTACTGACT ACCACCTGAA

 +1    G  L  Y  T  R  R  L  N  R  L  P  D  G  M  A  V  V  R  E  T  L  Q  R  N  T  S
 2401  GGCCTCTATA CCCGTCGCCT GAACCGGCTC CCTGATGGGA TGGCTGTGGT ACGGGAGACC CTGCAACGAA ATACCTCCCT
       CCGGAGATAT GGGCAGCGGA CTTGGCCGAG GGACTACCCT ACCGACACCA TGCCCTCTGG GACGTTGCTT TATGGAGGGA

 +1    G  L  G  D  A  D  S  W  D  D  S  S  S  V  S  S  G  I  S  D  T  I  D  N  L  S
 2481  GGGCCTCGGA GACGCTGACA GCTGGGACGA CAGCAGCTCC GTCAGCAGCG GCATCAGCGA CACCATAGAC AACCTCAGCA
       CCCGGAGCCT CTGCGACTGT CGACCCTGCT GTCGTCGAGG CAGTCGTCGC CGTAGTCGCT GTGGTATCTG TTGGAGTCGT
                                   97012801
```

11/25/97 01:00PM hh2UNC53_971120 Page 3

```
   +1 T D D I  N T S  S S I S  S Y A  H T P  A S S K  K N L  D V Q
2561 CTGATGACAT CAACACCAGC TCCTCCATCA GCTCTTATGC CAACACACCT GCCTCCTCTC GAAAAAACCT GGATGTGCAG
     GACTACTGTA GTTGTGGTCG AGGAGGTAGT CGAGAATACG GTTGTGTGGA CGGAGGAGAG CTTTTTTGGA CCTACACGTC

   +1 T D A E  K H S  Q V E  K N S L  W S G  D D V  K K S D  G G
2641 ACTGATGCTG AGAAGCACTC ACAGGTGGAG AAGAATTCCC TGTGGTCTGG TGATGATGTC AAGAAATCAG ACGGAGGCTC
     TGACTACGAC TCTTCGTGAG TGTCCACCTC TTCTTAAGGG ACACCAGACC ACTACTACAG TTCTTTAGTC TGCCTCCGAG

   +1  D S G  I K H E  P G S  K W R  R N P S  D V S  D E S  D K S
2721 AGACAGCGGC ATAAAAATGG AGCCAGGTTC CAAGTGGAGG CGGAATCCTT CTGATGTGTC TGACGAGTCC GACAAAAGCA
     TCTGTCGCCG TATTTTTACC TCGGTCCAAG GTTCACCTCC GCCTTAGGAA GACTACACAG ACTGCTCAGG CTGTTTTCGT

   +1 T S G K  K N P  V I S Q  T G S  W R R  G M T A  Q V G  I T M
2801 CGTCGGGAAA GAAGAATCCT GTCATCTCCC AGACAGGCTC ATGGCGGCGA GGCATGACAG CTCAGGTGGG CATCACCATG
     GCAGCCCTTT CTTCTTAGGA CAGTAGAGGG TCTGTCCGAG TACCGCCGCT CCGTACTGTC GAGTCCACCC GTAGTGGTAC

   +1 P P T F  P S A  P A G  A L F T  P G T  G K T  D D A K  V S
2881 [illegible] AAACTTCAAC CCCAAATAGA GCACTGAAAG CCCCAGGAAC TGAAAAAACA GACGACGCAA AGGTGTCTGA
     [illegible] ACCTTTTTGT CTGCTGCGTT TCCACAGACT

   +1 P G K  L S P K  A S Q  V K P  S P S D  A G N  S S G  D E S
2961 [illegible] CTTTCTCCTA AAGCTTCCCA GGTGAAGCCC TCTCCATCAG ATGCAGGAAA CAGCAGTGGT GACGAATCCA
     [illegible] AGGAGGTAGTC TACGTCCTTT GTCGTCACCA CTGCTTAGGT
                                               97012802

   +1 K K P L  P S S  S R T P  T A N  A N S  F G F K  K Q S  G S A
3041 AAAAGCCCCT CCCCAGCAGC TCTAGGACAC CTACTGCCAA TGCCAACAGC TTTGGCTTCA AGAAGCAGAG TGGTTCCGCC
     TTTTCGGGGA GGGGTCGTCG AGATCCTGTG GATGACGGTT ACGGTTGTCG AAACCGAAGT TCTTCGTCTC ACCAAGGCGG

   +1 A G L A  H I T  A S G  V T V T  S K S  A T L  G K I P  K S
3121 GCCGGCCTGG CCATGATCAC AGCCAGCGGG GTGACTGTCA CCAGCAGGTC AGCCACACTG GGCAAAATCC CAAAGTCATC
     CGGCCGGACC GGTACTAGTG TCGGTCGCCC CACTGACAGT GGTCGTCCAG TCGGTGTGAC CCGTTTTAGG GTTTCAGTAG

   +1  A L V  S H S A  G R K  S S H  D G A Q  N Q D  D G Y  L A L
3201 TGCACTCGTC AGTCGGTCTG CTGGTCGGAA GTCAAGTATG GATGGGGCTC AGAATCAGGA TGACGGGTAT CTAGCCCTAA
     ACGTGAGCAG TCAGCCAGAC GACCAGCCTT CAGTTCATAC CTACCCCGAG TCTTAGTCCT ACTGCCCATA GATCGGGATT

   +1 S S R T  N L Q  Y R S L  P R P  S K S  N S R N  G A G  N R S
3281 GCTCCCGGAC AAACCTTCAG TACCGGAGTT TGCCGAGGCC CAGTAAGTCC AACAGCCGGA ACGGGGCTGG GAACAGGTCT
     CGAGGGCCTG TTTGGAAGTC ATGGCCTCAA ACGGCTCCGG GTCATTCAGG TTGTCGGCCT TGCCCCGACC CTTGTCCAGA

   +1 S T S S  I D S  N I S  S K S A  G L P  V P K  L R E P  S K
3361 AGCACCAGCA GCATAGATTC CAACATTAGC AGCAAGTCCG CAGGCCTGCC AGTGCCCAAA CTGAGGGAGC CTTCCAAAAC
     TCGTGGTCGT CGTATCTAAG GTTGTAATCG TCGTTCAGGC GTCCGGACGG TCACGGGTTT GACTCCCTCG GAAGGTTTTG

   +1  A L G  S S L P  G L V  N Q T  D K E K  G I S  S D N  E S V
3441 AGGCCTAGGC AGCTCTCTAC CAGGTCTGGT CAACCAAACA GATAAGGAGA AAGGCATCTC ATCAGACAAC GAGAGTGTGG
     TCGGGATCCG TCGAGAGATG GTCCAGACCA GTTGGTTTGT CTATTCCTCT TTCCGTAGAG TAGTCTGTTG CTCTCACACC

   +1 A S C H  S V K  V N P A  A Q P  V S S  P A Q T  S L Q  P G A
3521 CTTCCTGTAA CTCAGTGAAA GTGAATCCGG CAGCCCAGCC TGTGTCCAGT CCGGCTCAGA CCAGTCTCCA GCCTGGAGCC
     GAAGGACATT GAGTCACTTT CACTTAGGCC GTCGGGTCGG ACACAGGTCA GGCCGAGTCT GGTCAGAGGT CGGACCTCGG

   +1 K Y P D  V A S  P T L  R R L F  G G K  P T K  Q V P I  A T
3601 AAGTACCCAG ATGTGGCCTC TCCCACACTC CGCAGACTCT TTGGTGGGAA GCCTACCAAG CAAGTGCCCA TCGCCACAGC
     TTCATGGGTC TACACCGGAG AGGGTGTGAG GCGTCTGAGA AACCACCCTT CGGATGGTTC GTTCACGGGT AGCGGTGTCG

   +1  E N M  K N S V  V I S  N P H  A T M T  Q Q G  N L D  S P S
3681 TGAAAACATG AAAAATTCGG TGGTCATCTC CAATCCTCAT GCCACCATGA CTCAGCAAGG TAACCTAGAC TCCCCGTCAG
     ACTTTTGTAC TTTTTAAGCC ACCAGTAGAG GTTAGGAGTA CGGTGGTACT GAGTCGTTCC ATTGGATCTG AGGGGCAGTC

   +1 G S G V  L S S  G S S S  P L Y  S K N  V D L N  Q S P  L A S
3761 GCAGTGGCGT CCTGAGCAGT GGGAGCAGCA GTCCTCTCTA CAGCAAGAAT GTGGACCTCA ACCAGTCTCC GCTAGCCTCC
     CGTCACCGCA GGACTCGTCA CCCTCGTCGT CAGGAGAGAT GTCGTTCTTA CACCTGGAGT TGGTCAGAGG CGATCGGAGG
```

11/25/97 01:00PM hh2UNC53_971120 Page 4

```
   +1 S P S S  A H S  A P S  N S L T  W G T  N A S  S S S A  V S
3841 AGCCCCAGCT CAGCCCACTC GGCCCCTTCC AACAGCCTCA CCTGGGGCAC CAACGCCAGC AGCTCCTCCG CAGTTAGCAA
     TCGGGGTCGA GTCGGGTGAG CCGGGGAAGG TTGTCGGAGT GGACCCCGTG GTTGCGGTCG TCGAGGAGGC GTCAATCGTT

   +1  D G L  G F Q S  V S S  L H T  S C E S  I D I  S L S  S G G
3921 GGATGGCCTG GGCTTTCAGT CTGTCAGCAG CCTCCACACC AGCTGTGAGT CCATCGACAT CTCCCTCAGC AGTGGAGGGG
     CCTACCGGAC CCGAAAGTCA GACAGTCGTC GGAGGTGTGG TCGACACTCA GGTAGCTGTA GAGGGAGTCG TCACCTCCCC

   +1 V P S H  N S S  T G L I  A S S  K D D  S L T P  F V R  T N S
4001 TCCCCAGCCA CAATTCTTCC ACTGGCCTCA TCGCCTCCTC CAAGGACGAC TCCTTGACTC CCTTTGTCAG AACTAACAGT
     AGGGGTCGGT GTTAAGAAGG TGACCGGAGT AGCGGAGGAG GTTCCTGCTG AGGAACTGAG GGAAACAGTC TTGATTGTCA

   +1  V K T  T L S E  S P L  S S P  A A S P  K F C  R S T  L P R
4081 GTGAAGACCA CACTGTCAGA AAGCCCTCTC TCTTCCCCTG CTGCTAGCCC TAAGTTCTGC AGAAGTACTC TGCCCAGGAA
     CACTTCTGGT GTGACAGTCT TTCGGGAGAG AGAAGGGGAC GACGATCGGG ATTCAAGACG TCTTCATGAG ACGGGTCCTT

   +1  Q D S  D P H L  D R N  T L P  K K G L  R Y T  P T S  Q L R
4161 ACAGGACAGT GACCCGCACC TTGATAGGAA CACTTTGCCT AAGAAAGGAC TCAGGTATAC TCCCACCTCC CAGCTTCGCA
     TGTCCTGTCA CTGGGCGTGG AACTATCCTT GTGAAACGGA TTCTTTCCTG AGTCCATATG AGGGTGGAGG GTCGAAGCGT
                                   97010708

   +1 T Q E D  A K E  W L R S  H S A  G G L  Q D T A  A N S  P F S
4241 CGCAAGAAGA TGCAAAAGAA TGGTTACGGT CCCATTCTGC AGGAGGCCTT CAGGACACCG CTGCCAATTC CCCCTTTTCC
     GCGTTCTTCT ACGTTTTCTT ACCAATGCCA GGGTAAGACG TCCTCCGGAA GTCCTGTGGC GACGGTTAAG GGGGAAAAGG
              97010709

   +1 S G S S  V T S  P S G  T R P N  F S Q  L A S  P T T V  T Q
4321 TCTGGCTCCA GCGTGACTTC TCCCTCCGGA ACAAGATTCA ACTTTTCCCA GCTTGCGAGT CCCACCACTG TCACCCAGAT
     AGACCGAGGT CGCACTGAAG AGGGAGGCCT TGTTCTAAGT TGAAAAGGGT CGAACGCTCA GGGTGGTGAC AGTGGGTCTA

   +1  S L S  N P T M  L R T  H S L  S N A D  G Q Y  D P Y  T D S
4401 GAGCTTGTCC AACCCGACCA TGCTGAGGAC TCACAGCCTC TCCAATGCTG ATGGGCAGTA TGATCCATAC ACTGACAGCC
     CTCGAACAGG TTGGGCTGGT ACGACTCCTG AGTGTCGGAG AGGTTACGAC TACCCGTCAT ACTAGGTATG TGACTGTCGG

   +1 R F R N  S S M  S L D  E K S R  T M S  R S G  S F R D  G F E
4481 GCTTCCGGAA TAGCTCCATG TCCCTGGATG AGAAGAGCAG AACCATGAGC CGTTCAGGCT CATTCCGGGA TGGGTTTGAA
     CGAAGGCCTT ATCGAGGTAC AGGGACCTAC TCTTCTCGTC TTGGTACTCG GCAAGTCCGA GTAAGGCCCT ACCCAAACTT

   +1  E V H  G S S L  S L V  S S T  L S V Y  S T P  E E K  C Q S
4561 GAAGTTCATG GATCCTCACT CTCCCTGGTT TCCAGCACAT TGTCAGTTTA TTCTACACCA GAAGAAAAAT GCCAGTCAGA
     CTTCAAGTAC CTAGGAGTGA GAGGGACCAA AGGTCGTGTA ACAGTCAAAT AAGATGTGGT CTTCTTTTTA CGGTCAGTCT

   +1  I R K  L R R E  L D A  S Q E  K V S A  L T T  Q L T  A N A
4641 GATTCGCAAG CTGCGGCGGG AACTGGATGC CTCCCAGGAG AAAGTTTCAG CTTTGACCAC CCAGCTGACA GCAAATGCTC
     CTAAGCGTTC GACGCCGCCC TTGACCTACG GAGGGTCCTC TTTCAAAGTC GAAACTGGTG GGTCGACTGT CGTTTACGAG

   +1 H L V A  A F E  Q S L G  N M T  I R L  Q S L T  M T A  E Q K
4721 ACCTTGTGGC TGCCTTTGAA CAGAGTCTTG GTAACATGAC AATCAGGCTC CAGAGTCTGA CCATGACAGC TGAGCAGAAG
     TGGAACACCG ACGGAAACTT GTCTCAGAAC CATTGTACTG TTAGTCCGAG GTCTCAGACT GGTACTGTCG ACTCGTCTTC

   +1 D S E L  N E L  R K T  I E L L  K K Q  N A A  A Q A A  I N
4801 GATTCAGAAC TGAATGAGTT AAGAAAAACC ATTGAGCTGC TAAAGAAACA GAACGCAGCT GCCCAGGCTG CCATTAATGG
     CTAAGTCTTG ACTTACTCAA TTCTTTTTGG TAACTCGACG ATTTCTTTGT CTTGCGTCGA CGGGTCCGAC GGTAATTACC

   +1  V I N  T P E L  N C K  G N G  T A Q S  A D L  R I R  R Q H
4881 AGTAATTAAC ACACCTGAGC TCAACTGCAA AGGAAACGGC ACTGCCCAGT CTGCAGACCT CCGCATCCGC AGGCAGCACT
     TCATTAATTG TGTGGACTCG AGTTGACGTT TCCTTTGCCG TGACGGGTCA GACGTCTGGA GGCGTAGGCG TCCGTCGTGA

   +1 S S D S  V S S  I N S A  T S H  S S V  G S N I  E S D  S K K
4961 CCTCAGACAG CGTCTCCAGC ATCAACAGTG CCACCAGCCA CTCCAGTGTG GGCAGCAACA TAGAGAGTGA CTCAAAGAAG
     GGAGTCTGTC GCAGAGGTCG TAGTTGTCAC GGTGGTCGGT GAGGTCACAC CCGTCGTTGT ATCTCTCACT GAGTTTCTTC

   +1 K K R K  N W L  R S S  F K Q A  F G K  K K S  P K S A  S S
5041 AAGAAGAGGA AGAACTGGTT ACGCAGCTCC TTCAAGCAAG CTTTCGGGAA GAAGAAGTCC CCAAAATCTG CGTCCTCTCA
     TTCTTCTCCT TCTTGACCAA TGCGTCGAGG AAGTTCGTTC GAAAGCCCTT CTTCTTCAGG GGTTTTAGAC GCAGGAGAGT
```

```
11/25/97 01:00PM        hh2UNC53_971120        Page 5

+1    V  I  I  L  E  E  T  T  D  S  S  L  P  S  S  P  K  L  P  H  N  G  S  T  G  S  T
5121  TTTAGATATT GAGGAGACGA CGGATTCTTC TTTACCTTCC TCACCAAAGT TACCCCACAA TGGGTCCACA GGTTCCACCC
      AAATCTATAA CTCCTCTGCT GCCTAAGAAG AAATGGAAGG AGTGGTTTCA ATGGGGTGTT ACCCAGGTGT CCAAGGTGGG

+1 F  L  L  K  H  S  H  S  H  S  L  I  S  E  C  H  D  S  E  A  E  T  V  H  Q  L  R
5201  CATTACTGAG GAATTCTCAC TCCAACTCTC TAATTTCCGA ATGCATGGAT AGTGAAGCTG AGACCGTCAT GCAGCTCCGA
      GTAATGACTC CTTAAGAGTG AGGTTGAGAG ATTAAAGGCT TACGTACCTA TCACTTCGAC TCTGGCAGTA CGTCGAGGCT

+1 H  E  L  K  D  K  E  M  K  L  T  D  I  H  L  E  A  L  S  S  A  H  Q  L  D  Q
5281  AATGAGTTAA GAGACAAGGA GATGAAGCTG ACGGATATCC ACTTAGAAGC TCTCAGTTCT GCCCACCAGC TGGACCAGCT
      TTACTCAATT CTCTGTTCCT CTACTTCGAC TGCCTATAGG TGAATCTTCG AGAGTCAAGA CGGGTGGTCG ACCTGGTCGA
                                               EST 46037-R2

+1    F  E  A  H  H  R  H  V  S  E  I  E  K  L  K  A  E  N  D  K  L  K  S  E  S  Q
5361  [illegible] ATGAACAGGA TGAGAGTGAA AATAGAGAAG CTGAAAGCTG AGAATGATCG GCTGAAGTCA GAGTCTCAAG
      [illegible] TACTTGTCCT ACTCTCACTT TTATCTCTTC GACTTTCGAC TCTTACTAGC CGACTTCAGT CTCAGAGTTC
                                                                     EST 46037-R1

+1 G  S  S  C  S  R  A  P  S  Q  V  S  I  S  A  S  P  R  Q  S  H  G  L  S  Q  H  S
5441  GGAGCAACTG CAGCCGGACT CCTTCCCAAG TGTCCATCTC TGCCTCCCCG AGGCAGTCCA TGGCCTCTCC CCAGCACAGC
      [illegible]

+1 L  H  L  F  E  S  T  S  L  D  H  L  L  D  D  T  G  E  P  S  A  K  K  E  G  G
5521  TTGAACTTTA CTGAGTCAAC CAGTCTGGAC ATGTTGCTGG ATGACACTGG TGAATTTGTG GCTCGGAAGG AAGGAGGTAG
      [illegible]

+1    H  V  F  I  V  V  S  F  Q  E  E  H  K  W  K  E  D  S  K  P  H  L  P  L  I  G
5601  GCATGTTAAG ATAGTTGTCA GCTTTCAGGA GGAAATGAAG TGGAAGGACG ATTCCAGACC ACACCTCTTT CTTATTGGCT
      CGTACAATTC TATCAACAGT CGAAAGTCCT CCTTTACTTC ACCTTCCTGC TAAGGTCTGG TGTGGAGAAA GAATAACCGA

+1 C  I  G  V  S  G  K  T  K  W  D  V  L  D  G  V  V  R  R  L  F  K  E  Y  I  I  H
5681  GCATTGGACT TAGTGGCAAG ACGAAGTGGG ATGTGCTGGA TGGGGTGGTT AGACGGCTGT TCAAAGAATA CATCATTCAT
      CGTAACCTGA ATCACCGTTC TGCTTCACCC TACACGACCT ACCCCACCAA TCTGCCGACA AGTTTCTTAT GTAGTAAGTA

+1    V  D  P  V  S  Q  L  G  L  H  S  D  S  V  L  G  Y  S  I  G  E  I  K  R  S  N
5761  GTGGACCCAG TGAGTCAGCT AGGGCTGAAT TCAGACAGCG TTCTTGGCTA CAGCATTGGA GAAATCAAGC GCAGCAACAC
      CACCTGGGTC ACTCAGTCGA TCCCGACTTA AGTCTGTCGC AAGAACCGAT GTCGTAACCT CTTTAGTTCG CGTCGTTGTG
                                EST-46037

+1    S  E  T  P  E  L  L  P  C  G  Y  L  V  G  E  N  T  T  I  S  V  T  V  K  G  L
5841  TTCCGAAACA CCGGAGCTGC TTCCTTGTGG CTATCTGGTT GGAGAGAACA CGACCATCTC AGTGACTGTG AAAGGGCTCG
      AAGGCTTTGT GGCCTCGACG AAGGAACACC GATAGACCAA CCTCTCTTGT GCTGGTAGAG TCACTGACAC TTTCCCGAGC

+1 A  E  H  S  L  D  S  L  V  F  E  S  L  I  P  K  P  I  L  Q  H  Y  V  S  L  L  I
5921  CAGAAAACAG CCTGGACTCA CTAGTGTTTG AGTCCTTGAT TCCCAAGCCC ATCCTGCAGC GCTACGTCTC CCTCCTGATA
      GTCTTTTGTC GGACCTGAGT GACCACAAAC TCAGGAACTA AGGGTTCGGG TAGGACGTCG CGATGCAGAG GGAGGACTAT

+1    E  H  R  K  I  I  L  S  G  P  S  G  T  G  K  T  Y  L  A  N  R  L  S  E  Y  I
6001  GAGCACCGTC GGATCATTCT CTCTGGCCCC AGCGGCACTG GGAAAACCTA CCTGGCCAAC CGGCTGTCTG AGTATATAGT
      CTCGTGGCAG CCTAGTAAGA GAGACCGGGG TCGCCGTGAC CCTTTTGGAT GGACCGGTTG GCCGACAGAC TCATATATCA

+1    L  R  E  G  R  E  L  T  D  G  V  I  A  T  F  N  V  D  H  K  S  S  K  E  L  R
6081  GCTTCGAGAG GGACGGGAGT TGACAGACGG GGTTATCGCC ACCTTTAACG TGGACCATAA GTCCAGCAAG GAATTGCGCC
      CGAAGCTCTC CCTGCCCTCA ACTGTCTGCC CCAATAGCGG TGGAAATTGC ACCTGGTATT CAGGTCGTTC CTTAACGCGG

+1 Q  Y  L  S  N  L  A  D  Q  C  N  S  E  N  N  A  V  D  H  P  L  V  I  I  L  D  N
6161  AGTACCTGTC CAACCTTGCT GACCAGTGCA ACAGTGAGAA CAATGCTGTG GACATGCCCC TCGTCATCAT CCTGGACAAC
      TCATGGACAG GTTGGAACGA CTGGTCACGT TGTCACTCTT GTTACGACAC CTGTACGGGG AGCAGTAGTA GGACCTGTTG
                                       EST46037-F1

+1 L  H  H  V  S  S  L  G  E  I  F  H  G  L  L  N  C  K  Y  H  K  C  P  Y  I  I
6241  CTACACCACG TCAGCTCTCT GGGCGAGATC TTCAATGGGC TGCTCAACTG CAAGTACCAC AAATGCCCTT ACATAATTGG
      GATGTGGTGC AGTCGAGAGA CCCGCTCTAG AAGTTACCCG ACGAGTTGAC GTTCATGGTG TTTACGGGAA TGTATTAACC
                                           EST46037-P2

+1    T  H  H  Q  A  T  S  S  T  P  N  L  Q  L  H  H  N  F  R  W  V  L  C  A  N  H
6321  CACAATGAAC CAGGCTACCT CTTCGACTCC CAACCTGCAG CTTCACCATA ACTTCAGATG GGTCCTTTGT GCCAACCACA
      GTGTTACTTG GTCCGATGGA GAAGCTGAGG GTTGGACGTC GAAGTGGTAT TGAAGTCTAC CCAGGAAACA CGGTTGGTGT
```

```
11/25/97 01:00PM        hh2UNC53_971120        Page 6

+1 T  E  P  V  K  G  F  L  G  R  F  L  R  R  K  L  M  E  T  E  I  S  G  R  V  R  N
6401  CGGAGCCTGT GAAGGGTTTC CTTGGCCGAT TCCTGAGGAG GAAGCTCATG GAAACAGAGA TCAGTGGGCG GGTGCGCAAT
      GCCTCGGACA CTTCCCAAAG GAACCGGCTA AGGACTCCTC CTTCGAGTAC CTTTGTCTCT AGTCACCCGC CCACGCGTTA

+1    M  E  L  V  K  I  I  D  W  I  P  K  V  W  H  H  L  N  R  F  L  E  A  H  S  S
6481  ATGGAGCTGG TAAAAATCAT TGACTGGATT CCCAAGGTCT GGCATCACCT CAACCGCTTC CTGGAGGCTC ACAGTTCCTC
      TACCTCGACC ATTTTTAGTA ACTGACCTAA GGGTTCCAGA CCGTAGTGGA GTTGGCGAAG GACCTCCGAG TGTCAAGGAG

+1    D  V  T  I  G  P  R  L  F  L  S  C  P  I  D  V  D  G  S  R  V  W  F  T  D  L
6561  GGACGTCACC ATCGGCCCCC GGCTCTTCCT GTCATGCCCC ATCGATGTGG ACGGCTCGAG AGTGTGGTTC ACCGACTTGT
      CCTGCAGTGG TAGCCGGGGG CCGAGAAGGA CAGTACGGGG TAGCTACACC TGCCGAGCTC TCACACCAAG TGGCTGAACA

+1 W  N  Y  S  I  I  P  Y  L  L  E  A  V  R  E  G  L  Q  L  Y  G  R  R  A  P  W  E
6641  GGAACTATTC CATTATCCCC TATCTCCTGG AAGCCGTCAG AGAAGGACTC CAGCTCTATG GAAGGCGCGC CCCCTGGGAG
      CCTTGATAAG GTAATAGGGG ATAGAGGACC TTCGGCAGTC TCTTCCTGAG GTCGAGATAC CTTCCGCGCG GGGGACCCTC

+1    D  P  A  K  N  V  N  D  T  Y  P  W  A  A  S  P  Q  Q  H  E  W  P  P  L  L  Q
6721  GATCCTGCCA AGTGGGTGAT GGACACATAT CCATGGGCAG CCAGCCCACA ACAGCACGAG TGGCCTCCCC TGCTGCAGTT
      CTAGGACGGT TCACCCACTA CCTGTGTATA GGTACCCGTC GGTCGGGTGT TGTCGTGCTC ACCGGAGGGG ACGACGTCAA

+1    R  P  E  D  V  G  F  D  G  Y  S  M  P  R  E  G  S  T  S  K  Q  M  P  P  S  D
6801  ACGGCCTGAG GATGTCGGCT TCGACGGCTA CTCCATGCCT CGGGAGGGAT CGACAAGCAA GCAGATGCCC CCCAGTGATG
      TGCCGGACTC CTACAGCCGA AGCTGCCGAT GAGGTACGGA GCCCTCCCTA GCTGTTCGTT CGTCTACGGG GGGTCACTAC

+1 A  E  G  D  P  L  M  N  M  L  M  R  L  Q  E  A  A  N  Y  S  S  P  Q  S  Y  D  S
6881  CTGAAGGTGA CCCGCTGATG AACATGCTGA TGAGGCTGCA GGAGGCAGCC AACTACTCCA GCCCCCAGAG CTATGACAGC
      GACTTCCACT GGGCGACTAC TTGTACGACT ACTCCGACGT CCTCCGTCGG TTGATGAGGT CGGGGGTCTC GATACTGTCG
                                    EST-923793

+1 D  S  N  S  N  S  H  H  E  D  I  L  D  S  S  L  E  S  T  L  *  Q  G  P  G  A
6961  GACTCCAACA GCAACAGCCA TCACGAAGAC ATCTTGGACT CCTCTTTGGA GTCCACTCTG TGACAGGGGC CCGGAGCCCA
      CTGAGGTTGT CGTTGTCGGT AGTGCTTCTG TAGAACCTGA GGAGAAACCT CAGGTGAGAC ACTGTCCCCG GGCCTCGGGT

+1    R  P  P  L  L  L  T  A  F  H  L  H  P  P  H  H  P  E  D  D  F  L  S  Q  P  P
7041  GCGCCCTCCT CTTCTCCTCA CCGCATTCCA CCTGCATCCC CCACATCACC CTGAAGATGA CTTCCTGAGC CAGCCCCCAG
      CGCGGGAGGA GAAGAGGAGT GGCGTAAGGT GGACGTAGGG GGTGTAGTGG GACTTCTACT GAAGGACTCG GTCGGGGGTC

+1 A  T  A  L  E  L  R  E  H  R  D  P  P  S  F  S  L  D  L  G  A  G  I  P  G  Q  L
7121  CCACAGCCTT ACAGCTCCGG GAACACCGAG ACCCCCCGTC CTTCAGCCTC GACCTGGGTG CAGGCATCCC GGGCCAGCTG
      GGTGTCGGAA TCTCGACGCC CTTGTGGCTC TGGGGGGCAG GAAGTCGGAG CTGGACCCAC GTCCGTAGGG CCCGGTCGAC

+1 P  A  D  R  F  L  P  Q  R  E  L  H  Y  L  L  L  Y  F  N  Y  C  F  A  L  L  L
7201  CCTGCGGACC GCTTCCTTCC ACAGCGAGAA CTGCACTACC TTCTGTTGTA CTTTAATTAT TGTTTTGCCT TGTTGCTGTG
      GGACGCCTGG CGAAGGAAGG TGTCGCTCTT GACGTGATGG AAGACAACAT GAAATTAATA ACAAAACGGA ACAACGACAC

+1    P  P  *  D  T  E  D  T  S  R  E  R  I  I  A  V  E  M  K  K  K  K  K  K  K  K
7281  ACCTCCCTAA GACACTGAAG ATACTTCTCG GGAAAGGATC ATCGCCGTTG AAATGAAAAA AAAAAAAAAA AAAAAAAAAA
      TGGAGGGATT CTGTGACTTC TATGAAGAGC CCTTTCCTAG TAGCGGCAAC TTTACTTTTT TTTTTTTTTT TTTTTTTTTT

+1 K  K  K  K
7361  AAAAAAAAAA AAA
      TTTTTTTTTT TTT
```

SEQUENCE ID NO 5

GAATTCCGGGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGG 100
N S G E E L T V S G S P A A G Q L O S N Q R O A N T L P K K G L R
not part of ...

TACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCC 200
Y Q L Q S O E E T K E R R H S H T I G G L P E S O O Q S E L P S P

CTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACAGCATCCC 300
P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S N S I P

CACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGGATCCTTC 400
T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R S G S F

CGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAAA 500
R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M Q S E Q
homology block A

TCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCA 600
I R K L R R E L E S S Q E K V A T L T S Q L S A N A N L V A A F E Q
homology block A

GAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGACTTTCTG 700
S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D F L
homology block A

AAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACTCCTCAG 800
K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q N S S
homology block A

ATAGCATCTCAAGCCTCAACAGCCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGGTCTATGA 900
D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S W V Y E
homology block B

GCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCA 1000
L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D S S
homology block B

GCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTCACCGAGG 1100
A P S S P K L Q H G S T E T A S P S I K S S T L S S V G T D V T E

GCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGA 1200
G P A H P A P - T R L F H A N E E E E P E K K E V S E L R S E L W E
homology block C

GAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTG 1300
K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L E V
homology block C

GACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCC 1400
D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L S S
homology block C
weak homology in hum1 vs hum2

CACGCCGCTCCCTAGGCCTGCGACTGACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAA 1500
P R R S L G L R L T H S F G P S L A D T D L S P M D G I S T C G P K
weak homology in hum1 vs hum2

GGAGGAAGTGACCCTCCGGCTGCTGGTGAGGATGCCCCCGCAGCACATCATCAAGGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTC 1500
E E V T L R L V R M P P Q H I I K G D L K Q Q E F F L G C S K V
homology block D
weak homology in hum1 vs hum2

AGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCA 1700
S G K V D W K M L D E A V F Q V F K D Y I S K M D P A S T L G L S
homology block D

CTGAGTCCATCCATGGCTACAGCCATTAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGGCGAGGTGTCAATAACATATC 1800
T E S I H G Y S H S H V K R V L D A E P P E M P P C R R G V N N I S
homology block D

AGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTG 1900
V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L L L
homology block E - pred nucleotide 8D

AAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGG 2000
K H R R L V L S S P S G T G K T Y L T N R L A E Y L V E R S G R E
homology block E - pred nucleotide 8D

TCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAAC 2100
V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D R E T
homology block E - pred nucleotide 8D

AGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCAT 2200
G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C K Y H
homology block E - pred nucleotide 8D

AAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACG 2300
K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S N N
homology block E - pred nucleotide 8D

TGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCT 2400
V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L R V L
homology block E - pred nucleotide 8D

CGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCC 2500
D V V P K L V V H L H T F L E K H S T S D F L I G P C F F L S C P
homology block E - pred nucleotide 8D

ATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATG 2600
I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G I K V H
homology block E - pred nucleotide 8D

GACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCC 2700
G Q K A A V E D P V E V V R D T L P V P S A Q Q D Q S K L Y H L P P
homology block E - pred nucleotide 8D

ACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATG 2800
P T V G P H S I A S P P E D R T V K D S T P S S L D S D P L M A M
homology block E - pred nucleotide 8D

CTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTCGGCAATC 2900
L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L
homology block E - pred nucleotide 8D

ACTGTCACCCCCCGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGA 3000
3' untranslated trailer

ACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTG 3100
3' untranslated trailer

TGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGG 3200
3' untranslated trailer

CTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGGAGATCTG 3300
3' untranslated trailer

GGGGAGGCAGGAAGCTCCTCAGATTTTTCTCACAGACCCTTCCCAATTCCATCACCACTGCCAACAACTCCTCCCCCAGAGATCTGGCTGGAGCCCAGAAA 3400
3' untranslated trailer

AAGAAGCATGTGGTTTAAAAAATGTTTAAATTAATCTGTAAAAGGTAAAATGAAAAAACAAAAACAAGCAAACAAACAAAAAACAATGGAAAAGATGAA 3500
3' untranslated trailer

GCTGGAGAGAGAGGAAACCAGTTTCCCAAGCTACAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACCTGAGAAG 3600
3' untranslated trailer

TCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAGCTTGGAC 3700
3' untranslated trailer

TTAACC 3706

SEQUENCE ID No. 6

Tuesday, 18 November 1997 13:57
lig 54 pLM1 (1 > 8285) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

```
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
CACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTA 100
G G T F R G N V R G T P I C L F F . I H S N M Y P L M R Q . P . . M

GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTG 200
L Q . Y . K R K S M S I Q H F R V A L I P F F A A F C L P V F A H

CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
GGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAA 300
P E T L V K V K D A E D Q L G A R V G Y I E L D L N S G K I L E S

TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
AAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTCGTTGAGCCAGC 400
F R P E E R F P M M S T F K V L L C G A V L S R I D A G Q E Q L G R

CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
GGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGG 500
R I H Y S Q N D L V E Y S P V T E K H L T D G M T V R E L C S A A

ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA
TATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATT 600
I T M S D N T A A N L L L T T I G G P K E L T A F L H N M G D H V

CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT
GAGCGGAACTAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGA 700
T R L D R W E P E L N E A I P N D E R D T T M P V A M A T T L R K L

ATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCG
TAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGC 800
L T G E L L T L A S R Q Q L I D W M E A D K V A G P L L R S A L P

GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG
CGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATC 900
A G W F I A D K S G A G E R G S R G I I A A L G P D G K P S R I V

TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCA
AATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGT 1000
V I Y T T G S Q A T M D E R N R Q I A E I G A S L I K H W . L S D Q

AGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCT
TCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGA 1100
V Y S Y I L . I D L K L H F . F K R I . V K I L F D N L M T K I P

TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA
ATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTT 1200
. R E F S F H . A S D P V E K I K G S S . D P F F L R V I C C L Q

CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAA
GTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTT 1300
T K K P P L P A V V C L P D Q E L P T L F P K V T G F S R A Q I P N

TACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
ATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGA 1400
T V L L V . P . L G H H F K N S V A P P T Y L A L L I L L P V A A
```

Tuesday, 18 November 1997 13:57 Page 1

fig 54 pLM1 (1 > 8285) Site and Sequence

```
GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC 1500
CGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCG
 A  S  G  D  K  S  C  L  T  G  L  D  S  R  R  .  L  P  D  K  A  Q  R  S  G  .  T  G  G  S  C  T  Q

CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC 1600
GGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGG
 P  S  L  E  R  T  T  Y  T  E  L  R  Y  L  Q  R  E  L  .  E  S  A  T  L  P  E  G  R  K  A  D  R  Y  P

GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTT 1700
CCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAA
 V  S  G  R  V  G  T  G  E  R  T  R  E  L  P  G  G  N  A  V  Y  L  Y  S  P  V  G  F  R  H  L  .  L

GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG 1800
CTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAAC
 E  R  R  F  L  .  C  S  S  G  G  R  S  L  V  K  N  A  S  N  A  A  F  L  R  F  L  A  F  C  V  P  F

CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCG 1900
GAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGC
 A  H  M  F  F  P  A  L  S  P  D  S  V  D  N  R  I  T  A  F  E  .  A  D  T  A  R  R  S  R  T  T  E  R

CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTT 2000
GTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAA
 S  E  S  V  S  E  E  A  E  E  R  P  I  R  K  P  P  L  P  A  R  V  P  I  H  .  C  S  V  H  D  R  F

CCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT 2100
GGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACA
 P  D  W  K  A  G  S  E  R  N  A  I  N  V  S  .  L  T  H  .  A  P  Q  A  L  H  F  M  L  P  A  R  M

TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCT 2200
ACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATACTGGTACTAATGCGGTTCGCGCGTTAATTGGGAGTGATTTCCCTTGTTTTCGA
 L  C  G  I  V  S  G  .  Q  F  H  T  G  N  S  Y  D  H  D  Y  A  K  R  A  I  N  P  H  .  R  E  Q  K  L

GGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAG 2300
CCCATGGCCCGGGGGGGAGCTCCAGCTGCCATAGCTATTCGAACTATAGCTTAAGGACGTCGGGGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTC
                                                                      insert pLM1
                                                                            ORF pLM1
 G  T  G  P  P  L  E  V  D  G  I  D  K  L  D  I  E  F  L  Q  P  L  L  F  S  Q  M  L  D  P  E  S  Q

AGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGA 2400
TCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCGGACGCTCCCAGGGTCCACTGAGTGTCGAGGGACCTCT
                                          insert pLM1
                                           ORF pLM1
 R  K  R  T  V  Q  N  V  L  D  L  R  Q  N  L  E  E  T  M  S  S  L  R  G  S  Q  V  T  H  S  S  L  E

TGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCGCG 2500
ACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAGACAGTACCGCGATACCGGTCAGGTCAGGCGC
                                          insert pLM1
                                           ORF pLM1
 M  T  C  Y  D  S  D  D  A  N  P  R  S  V  S  S  L  S  N  R  S  S  P  L  S  V  R  Y  G  Q  S  S  P  R
```

GCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTACATGCACGGCGAACGGGCCCACTACTCCCACACC
CGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGG 2600

insert pLM1

ORF pLM1

L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H T

ATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGA
TACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTACT 2700

insert pLM1

ORF pLM1

M P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y M

GCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCTC
CGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTACTTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAG 2800

insert pLM1

ORF pLM1

S D S D L M G K T M T E D D D I T T G W D E S S S I S S G L S D A S

AGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGCAGGAACTCAACAATAGTGCTACGC
TCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACGAAGAGCGTCCTTGAGTTGTTATCACGATGCG 2900

insert pLM1

ORF pLM1

D N L S S E E F N A S S S _ N S L P S T P T A S R R N S T I V L R

ACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCCTAAAAAACTGGAGTACGACAGTGGTAGCC
TGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTCTTTCGGGGATTTTTTGACCTCATGCTGTCACCATCGG 3000

insert pLM1

ORF pLM1

T D S E K R S L A E S G L S W F S E S E E K A P K K L E Y D S G S

TGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGCCCTGAGAGCTGTGATGATTCATCCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGG
ACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCGGGACTCTCGACACTACTAAGTAGGTTCCCACCTCTTGACTTTTTCGGGTAGTCGGACCC 3100

insert pLM1

ORF pLM1

L K M E P G T S K W R R E R P E S C D D S S K G G E L K K P I S L G

CCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCACACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAACCT
GGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCGGGTCTCACGGGAGTTTCAGCGTCCGTTTGGA 3200

insert pLM1

ORF pLM1

H P G S L K K G K T P P V A V T S P I T H T A Q S A L K V A G K P

GAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATAGTGGGCTCCAACGCTCCTCCTCTGATGCTGGTCGGGACCGCCTGAGTGATGCTAAGA
CTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATCACCCGAGGTTGCGAGGAGGAGACTACGACCAGCCCTGGCGGACTCACTACGATTCT 3300

insert pLM1

ORF pLM1

E G K A T D K G K L A V K N S G L Q R S S S D A G R D R L S D A K

Tuesday, 18 November 1997 13:57 Page 4
fig 54 pLM1 (1 > 8285) Site and Sequence

AGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTCCTGCCACAGGCACAGCCACTGTCATGCAAACTGGTGG 3400
TCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAGGACGGTGTCCGTGTCGGTGACAGTACGTTTGACCACC

insert pLM1

ORF pLM1

A P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G G

TTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCA 3500
AAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCGGT

insert pLM1

ORF pLM1

S A T L S K I Q K S S G I P V K P V N G R K T S L D V S N S A E P

GGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGTTCTATGAGCGTGACCGGCGGGCGGGGTGGAC 3600
CCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCAAGATACTCGCACTGGCCGCCCGCCCCACCTG

insert pLM1

ORF pLM1

G F L A P G A R S N I Q Y R S L P R P A K S S S M S V T G G R G G

CTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGG 3700
GAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGGTCTGACTTCCTCGGATGGTTCCATCGGTCACC

insert pLM1

ORF pLM1

P R P V S S S I D P S L L S T K Q G G L T P S R L K E P T K V A S G

GCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGAGT 3800
CGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCGTCACCGGAACCTGAGTCTGTTGTAGAGGAACTTCTCA

insert pLM1

ORF pLM1

R T T P A P V N Q T D R E K E K A K A K A V A L D S D N I S L K S

ATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCAACCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGA 3900
TAACCGAGGGGTCTCTCATGAGGGTTCTTGGTTCGTTCGGTTGGGTGTCGGTGGTTCGACCGTCTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTTCT

insert pLM1

ORF pLM1

I G S P E S T P K N Q A S Q P T A T K L A E L P P T P L R A T A K

GCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAAGGTCCC 4000
CGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATGGTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGG

insert pLM1

ORF pLM1

S F V K P P S L A N L D K V N S N S L D L P S S S D T T H A S K V P

AGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGCACCCATCCTCAATATTAACTCAGCCAGCTTCTCC 4100
TCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGAGG

insert pLM1

ORF pLM1

D L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F S

Tuesday, 18 November 1997 13:57 Page 5
fig 54 pLM1 (1 > 8285) Site and Sequence

CAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGC 4200
GTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCGGACGTGTCCTCGTACCTCAGGGAGGTCTACG
insert pLM1
ORF pLM1
Q G L E L M S G F S V P K E T R M Y P K L S G L H R S M E S L Q M

CAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTCCCACAGAAGAAGAGACGGAAGAGCTGACTTG 4300
GTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGTCTTCTTCTCTGCCTTCTCGACTGAAC
insert pLM1
ORF pLM1
P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T W

GAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAG 4400
CTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCTCCTC
insert pLM1
ORF pLM1
S G S P R A G Q L D S N Q R D R N T L P K K G L R Y Q L Q S Q E E

ACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAGTG 4500
TGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCAC
insert pLM1
ORF pLM1
T K E R R H S H T I G G L P E S D D Q S E L P S P P A L P M S L S

CAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACAGCATCCCCACCCACGAGGCGGCCTTCGAGCT 4600
GTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTCGA
insert pLM1
ORF pLM1
A K G Q L T N I V S P T A A T T P R I T R S N S I P T H E A A F E L

GTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGGATCCTTCCGAGACCCCACGGACGATGTTCAC 4700
CATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCCTAGGAAGGCTCTGGGGTGCCTGCTACAAGTG
insert pLM1
ORF pLM1
Y S G S Q M G S T L S L A E R P K G M I R S G S F R D P T D D V H

GGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAAATCCGGAAGCTTCGTAGGGAACTGG 4800
CCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACC
insert pLM1
ORF pLM1
G S V L S L A S S A S S T Y S S A E E R M Q S E Q I R K L R R E L

AATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCG 4900
TTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAAAACTCGTCTCGGACCACTTATACTGTAGGGC
insert pLM1
ORF pLM1
E S S Q E K V A T L T S Q L S A N A N L V A A F E Q S L V N M T S R

Tuesday, 18 November 1997 13:57 Page 6
fig 54 pLM1 (1 > 8285) Site and Sequence

```
CCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAG 5000
GGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTTCTTGAGACTCCGGGTC
                              insert pLM1
                              ORF pLM1
 L  R  H  L  A  E  T  A  E  E  K  D  T  E  L  L  D  L  R  E  T  I  D  F  L  K  K  K  N  S  E  A  Q

GCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCA 5100
CGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCTGTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGT
                              insert pLM1
                              ORF pLM1
 A  V  I  Q  G  A  L  N  A  S  E  T  T  P  K  E  L  R  I  K  R  Q  N  S  S  D  S  I  S  S  L  N  S

TCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGC 5200
AGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCG
                              insert pLM1
                              ORF pLM1
 I  T  S  H  S  S  I  G  S  S  K  D  A  D  A  K  K  K  K  K  K  S  V  V  Y  E  L  R  S  S  F  N  K  A

GTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCCAAACTACAG 5300
CAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTC
                              insert pLM1
                              ORF pLM1
 F  S  I  K  K  G  P  K  S  A  S  S  Y  S  D  I  E  E  I  A  T  P  D  S  S  A  P  S  S  P  K  L  Q

CATGGTTCCACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCACA 5400
GTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGTGT
                              insert pLM1
                              ORF pLM1
 H  G  S  T  E  T  A  S  P  S  I  K  S  S  T  L  S  S  V  G  T  D  V  T  E  G  P  A  H  P  A  P  H

CTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACAT 5500
GATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCCTTTACTTCGAATGTCTGTA
                              insert pLM1
                              ORF pLM1
 T  R  L  F  H  A  N  E  E  E  E  P  E  K  K  E  V  S  E  L  R  S  E  L  V  E  K  E  M  K  L  T  D  I

CCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGAC 5600
GGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTG
                              insert pLM1
                              ORF pLM1
 R  L  E  A  L  N  S  A  H  Q  L  D  Q  L  R  E  T  M  H  N  M  Q  L  E  V  D  L  L  K  A  E  N  D

[illegible]ACTGAAGGTAGCCCAGGCCCCTCATCAGGCTCCACTCCAGGGGAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCAC 5700
[illegible]TGACTTCCATCGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCCTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTG
                              insert pLM1
                              ORF pLM1
 [illegible]  L  K  V  A  P  G  P  S  S  G  S  T  P  G  E  V  P  G  S  S  A  L  S  S  P  R  R  S  L  G  L  A
```

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence Page 1

TCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGT
AGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACTGGGAGGCCCACCA 5800
insert pLM1
ORF pLM1
L T H S F G P S L A D T D L S P M D G I S T C G P K E E V T L R V V

GGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATG
CCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCTAC 5900
insert pLM1
ORF pLM1
V R M P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D W K M

CTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCA
GACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGT 6000
insert pLM1
ORF pLM1
L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I H G Y S

TCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGA
AGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCT 6100
insert pLM1
ORF pLM1
I S H V K R V L D A E P P E M P P C R R G V N N I S V S L K G L K E

GAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCG
CTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGC 6200
insert pLM1
ORF pLM1
K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S

GGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCT
CCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGA 6300
insert pLM1
ORF pLM1
G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T

TCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGT
AGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCA 6400
insert pLM1
ORF pLM1
F N M H Q Q S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L V

GATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTCAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACC
CTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCAGTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCATGG 6500
insert pLM1
ORF pLM1
I L L D D L S E A G S I S Q L V N G A L T C K Y H K C P Y I I G T

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 9

```
ACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGG 6600
TGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACC
                                        insert pLM1
                                        ORF pLM1
 T  N  Q  P  V  K  M  T  P  N  H  G  L  H  L  S  F  R  M  L  T  F  S  N  N  V  E  P  A  N  G  F  L

TTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTA 6700
AAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACCAT
                                        insert pLM1
                                        ORF pLM1
V  R  Y  L  R  R  K  L  V  E  S  D  S  D  I  N  A  N  K  E  E  L  L  R  V  L  D  W  V  P  K  L  W  Y

TCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACC 6800
AGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGG
                                        insert pLM1
                                        ORF pLM1
 H  L  H  T  F  L  E  K  H  S  T  S  D  F  L  I  G  P  C  F  F  L  S  C  P  I  G  I  E  D  F  R  T

TGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACC 6900
ACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGG
                                        insert pLM1
                                        ORF pLM1
 W  F  I  D  L  W  N  N  S  I  I  P  Y  L  Q  E  G  A  K  D  G  I  K  V  H  G  Q  K  A  A  W  E  D

CAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCAT 7000
GTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTA
                                        insert pLM1
                                        ORF pLM1
P  V  E  W  V  R  D  T  L  P  W  P  S  A  Q  Q  D  Q  S  K  L  Y  H  L  P  P  P  T  V  G  P  H  S  I

TGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCC 7100
ACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGG
                                        insert pLM1
                                        ORF pLM1
 A  S  P  P  E  D  R  T  V  K  D  S  T  P  S  S  L  D  S  D  P  L  M  A  M  L  L  K  L  Q  E  A  A

AACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAAC 7200
TTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTG
                                        insert pLM1
                  ORF pLM1
N  Y  I  E  S  P  D  R  E  T  I  L  D  P  N  L  Q  A  T  L  .  G  F  G  N  H  C  H  P  R  T  A  E

GCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAG 7300
CGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTTTC
                                        insert pLM1
A  W  H  Q  L  S  .  L  L  L  S  P  L  L  F  Q  S  T  G  S  P  A  P  G  G  E  Q  E  G  G  G  D  E  R
```

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence Page 1

AGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGC
TCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGACCG 7400
insert pLM1
G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T G

CTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAA
GAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTTTT 7500
insert pLM1
L L . . L V G K D D S G S F P . L L V S I T N S V A F V G G V Q K

CATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGGCTGAACTTGCTCAAAAGAAGCCGAATTCCAGCACACTGGCGGCCGTTACT
GTAGTTTTGTGACGTCGTCAAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGTTTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAATGA 7600
insert pLM1
T S K H C S S S P E F S L D L T R L N L L K R S R I P A H V R P L L

AGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGG
TCAAGATCTCGCCGGCGGTGGCGCCACCTCGAGGTTAAGCGGGATATCACTCAGCATAATGCGCGCGAGTGACCGGCAGCAAAATGTTGCAGCACTGACC 7700
V L E R P P P R V S S N S P Y S E S Y Y A R S L A V V L Q R R D V

GAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAAC
CTTTTGGGACCGCAATGGGTTGAATTAGCGGAACGTCGTGTAGGGGGAAAGCGGTCGACCGCATTATCGCTTCTCCGGGCGTGGCTAGCGGGAAGGGTTG 7800
E N P G V T Q L N R L A A H P P F A S V R N S E E A R T D R P S Q

AGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
TCAACGCGTCGGACTTACCGCTTACCTGCGCGGGACATCGCCGCGTAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTC 7900
Q L R S L N G E V D A P C S G A L S A A G V V V T R S V T A T L A S

CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTC
GCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCCGAGGGAAATCCCAAG 8000
A L A P A P F A F F P S F L A T F A G F P R Q A L N R G L P L G F

CGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGA
GCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACT 8100
R F S A L R H L D P K K L D . G D G S R S G P S P . . T V F R P L

CGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCC
GCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACCTTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACGG 8200
T L E S T F F N S G L L F Q T G T T L N P I S V Y S F D L . G I L P

GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAG
CTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATTGCGCTTAAAATTGTTTTATAATTGCGAATGTTAAATC 8285
I S A Y V L K N E L I . Q K F N A N F N K I L T L T I .

SEQUENCE ID

Tuesday, 18 November 1997 11:48 Page 1
fig 34 pLM4 (1 > 10070) Site and Sequence
Enzymes : 100 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

```
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC  100
                                          pCMV
 .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  V  P  A  V  L  T

CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA  200
                                          pCMV
 A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S  N  A  N  R  D  F  P  L  T  S  M  G  G  V  F  T  V

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT  300
                                          pCMV
 N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A  P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V

CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT  400
                                          pCMV
 H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H  R  Y  Y  H  G  D  A  V  L  A  V  H  Q  V  A  V

TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT  500
                                          pCMV
 I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q  V  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .

ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT  600
                                  pCMV
 Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G  G  L  Y  K  Q  S  V  F  S  E  P  S  D  P  L  A  L

CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC  700
                                          EGFP
 P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG  800
                                          EGFP
 V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC  900
                                          EGFP
 T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 1000

EGFP

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT 1100

EGFP

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 1200

EGFP

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT 1300

EGFP

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGATAAGCTTGATATCGAATTCCTGCAGCC
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGAGCTCGAGTTCGAAGCTTAAGACGTCAGCTATTCGAACTATAGCTTAAGGACGTCGG 1400

EGFP

I T L G M D E L Y K S G L R S R A Q A S N S A V D K L D I E F L Q P

CCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGC
GGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCG 1500

insert pLM1

ORF pLM1

L L F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S

CTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGT
GACGCTCCCAGGGTCCACTGAGTGTCGAGGGACCTCTACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCA 1600

insert pLM1

ORF pLM1

L R G S Q V T H S S L E M T C Y D S D D A N P R S V S S L S N R S

CCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTG
GGGGAGACAGTACCGCGATACCGGTCAGGTCAGGCGCCGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGAC 1700

insert pLM1

ORF pLM1

S P L S W R Y G Q S S P R L Q A G D A P S V G G S C R S E G T P A W

Tuesday, 18 November 1997 11:48 Page 3

fig 34 pLM4 (1 > 10070) Site and Sequence

```
GTACATGCACGGCGAACGGGCCCACTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTG
CATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGAC 1800
                                            insert pLM1
                                            ORF pLM1
 Y  M  H  G  E  R  A  H  Y  S  H  T  M  P  M  R  S  P  S  K  L  S  H  I  S  R  L  E  L  V  E  S  L

GACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATG
CTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTAC 1900
                                            insert pLM1
                                            ORF pLM1
 D  S  D  E  V  D  L  K  S  G  Y  M  S  D  S  D  L  M  G  K  T  M  T  E  D  D  D  I  T  T  G  W  D

AAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCC
TTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGG 2000
                                            insert pLM1
                                            ORF pLM1
E  S  S  S  I  S  S  G  L  S  D  A  S  D  N  L  S  S  E  E  F  N  A  S  S  S  L  N  S  L  P  S  T  P

CACTGCTTCTCGCAGGAACTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAG
GTGACGAAGAGCGTCCTTGAGTTGTTATCACGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTC 2100
                                            insert pLM1
                                            ORF pLM1
 T  A  S  R  R  N  S  T  I  V  L  R  T  D  S  E  K  R  S  L  A  E  S  G  L  S  W  F  S  E  S  E  E

AAAGCCCCTAAAAAACTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCAT
TTTCGGGGATTTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTA 2200
                                            insert pLM1
                                            ORF pLM1
K  A  P  K  K  L  E  Y  D  S  G  S  L  K  M  E  P  G  T  S  K  W  R  R  E  R  P  E  S  C  D  D  S

CCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCA
GGTTCCCACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGT 2300
                                            insert pLM1
                                            ORF pLM1
S  K  G  G  E  L  K  K  P  I  S  L  G  H  P  G  S  L  K  K  G  K  T  P  P  V  A  V  T  S  P  I  T  H
```

CACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCC 2400
GTGTCGGGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGG

insert pLM1

ORF pLM1

T A Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S

TCTGATGCTGGTCGGGACCGCCTGAGTGATGCTAAGAAGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTC 2500
AGACTACGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAG

insert pLM1

ORF pLM1

S D A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P

CTGCCACAGGCACAGCCACTGTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCG 2600
GACGGTGTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGC

insert pLM1

ORF pLM1

P A T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R

CAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAG 2700
GTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCGGTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTC

insert pLM1

ORF pLM1

K T S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A I

TCAAGTTCTATGAGCGTGACCGGCGGGCGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGC 2800
AGTTCAAGATACTCGCACTGGCCGCCCGCCCCACCTGGAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCG

insert pLM1

ORF pLM1

S S S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T

CTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGGCCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGC 2900
GAAGGTCTGACTTCCTCGGATGGTTCCATCGGTCACCGGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCG

insert pLM1

ORF pLM1

F S R L K E P T K V A S G R T T P A P V N Q T D R E K E K A K A K A

AGTGGCCTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCA 3000
TCACCGGAACCTGAGTCTGTTGTAGAGGAACTTCTCATAACCGAGGGGTCTCTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGT

insert pLM1

ORF pLM1

V A L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A

GAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTAC 3100
CTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATG

insert pLM1

ORF pLM1

E L P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L

CATCATCCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCC 3200
GTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGG

insert pLM1

ORF pLM1

P S S S D T T H A S K V P D L H A T S S A S G G P L P S C F T P S P

GGCACCCATCCTCAATATTAACTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTC 3300
CCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGAGGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAG

insert pLM1

ORF pLM1

A P I L N I N S A S F S Q G L E L M S G F S V P K E T R M Y P K L

TCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTG 3400
AGTCCGGACGTGTCCTCGTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGAC

insert pLM1

ORF pLM1

S G L H R S M E S L Q M P M S L P S A F P S S T P V P T P P A P P

CTGCTCCCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAA 3500
GACGAGGGTGTCTTCTTCTCTGCCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTT

insert pLM1

ORF pLM1

A A P T E E E T E E L T W S G S P R A G Q L D S N Q R D R N T L P K

Tuesday, 18 November 1997 11:48

fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

```
GAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAG
CTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCTCCTCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTC 3600
                                         insert pLM1
                                         ORF pLM1
 K  G  L  R  Y  Q  L  Q  S  Q  E  E  T  K  E  R  R  H  S  H  T  I  G  G  L  P  E  S  D  D  Q  S  E

CTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCT
GACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGA 3700
                                         insert pLM1
                                         ORF pLM1
L  P  S  P  P  A  L  P  M  S  L  S  A  K  G  Q  L  T  N  I  V  S  P  T  A  A  T  T  P  R  I  T  R

CCAACAGCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCG
GGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGC 3800
                                         insert pLM1
                                         ORF pLM1
S  N  S  I  P  T  H  E  A  A  F  E  L  Y  S  G  S  Q  M  G  S  T  L  S  L  A  E  R  P  K  G  M  I  R

GTCAGGATCCTTCCGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATG
CAGTCCTAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTAC 3900
                                         insert pLM1
                                         ORF pLM1
 S  G  S  F  R  D  P  T  D  D  V  H  G  S  V  L  S  L  A  S  S  A  S  S  T  Y  S  S  A  E  E  R  M

CAATCTGAGCAAATCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGG
GTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACC 4000
                                         insert pLM1
                                         ORF pLM1
Q  S  E  Q  I  R  K  L  R  R  E  L  E  S  S  Q  E  K  V  A  T  L  T  S  Q  L  S  A  N  A  N  L  V

CTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAAC
GACGAAAACTCGTCTCGGACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTG 4100
                                         insert pLM1
                                         ORF pLM1
A  A  F  E  Q  S  L  V  N  M  T  S  R  L  R  H  L  A  E  T  A  E  E  K  D  T  E  L  L  D  L  R  E  T
```

Tuesday, 18 November 1997 11:48 Page 1
fig 34 pLM4 (1 > 10070) Site and Sequence

```
CATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGA
-------------------------------------------------------------------------------------------------------+ 4200
GTATCTGAAAGACTTCTTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCT

insert pLM1

ORF pLM1
 I  D  F  L  K  K  K  N  S  E  A  Q  A  V  I  Q  G  A  L  N  A  S  E  T  T  P  K  E  L  R  I  K  R

CAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGA
-------------------------------------------------------------------------------------------------------+ 4300
GTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCT

insert pLM1

ORF pLM1
 Q  N  S  S  D  S  I  S  S  L  N  S  I  T  S  H  S  S  I  G  S  S  K  D  A  D  A  K  K  K  K  K  K

GTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTAC
-------------------------------------------------------------------------------------------------------+ 4400
CAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATG

insert pLM1

ORF pLM1
S  V  V  Y  E  L  R  S  S  F  N  K  A  F  S  I  K  K  G  P  K  S  A  S  S  Y  S  D  I  E  E  I  A  T

ACCCGACTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCCACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACT
-------------------------------------------------------------------------------------------------------+ 4500
TGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGA

insert pLM1

ORF pLM1
 P  D  S  S  A  P  S  S  P  K  L  Q  H  G  S  T  E  T  A  S  P  S  I  K  S  S  T  L  S  S  V  G  T

GATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCT
-------------------------------------------------------------------------------------------------------+ 4600
CTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGA

insert pLM1

ORF pLM1
D  V  T  E  G  P  A  H  P  A  P  H  T  R  L  F  H  A  N  E  E  E  E  P  E  K  K  E  V  S  E  L  R

CTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACAT
-------------------------------------------------------------------------------------------------------+ 4700
GACTCGATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTA

insert pLM1

ORF pLM1
S  E  L  W  E  K  E  M  K  L  T  D  I  R  L  E  A  L  N  S  A  H  Q  L  D  Q  L  R  E  T  M  H  N  M
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 9

GCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCT
CGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGA 4800

insert pLM1

ORF pLM1

Q L E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S

GCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTA
CGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCAT 4900

insert pLM1

ORF pLM1

A L S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S

CTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGG
GAACACCAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCC 5000

insert pLM1

ORF pLM1

T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G

CTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACC
GACATCGTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGG 5100

insert pLM1

ORF pLM1

C S K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T

CTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTG
GACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCAC 5200

insert pLM1

ORF pLM1

L G L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G

TCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACAT
AGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTA 5300

insert pLM1

ORF pLM1

V N N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 9

```
AAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGC
TTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCG  5400
                                        insert pLM1
                                        ORF pLM1
 S  L  L  L  K  H  R  R  L  V  L  S  G  P  S  G  T  G  K  T  Y  L  T  N  R  L  A  E  Y  L  V  E  R

TCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGA
AGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCT  5500
                                        insert pLM1
                                        ORF pLM1
 S  G  R  E  V  T  E  G  I  V  S  T  F  N  M  H  Q  Q  S  C  K  D  L  Q  L  Y  L  S  N  L  A  N  Q

TAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCAC
ATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTG  5600
                                        insert pLM1
                                        ORF pLM1
 I  D  R  E  T  G  I  G  D  V  P  L  V  I  L  L  D  D  L  S  E  A  G  S  I  S  E  L  V  N  G  A  L  T

CTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACC
GACGTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGG  5700
                                        insert pLM1
                                        ORF pLM1
 C  K  Y  H  K  C  P  Y  I  I  G  T  T  N  Q  P  V  K  M  T  P  N  H  G  L  H  L  S  F  R  M  L  T

TTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGC
AAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCG  5800
                                        insert pLM1
                                        ORF pLM1
 F  S  N  N  V  E  P  A  N  G  F  L  V  R  Y  L  R  R  K  L  V  E  S  D  S  D  I  N  A  N  K  E  E

TGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTT
ACGAAGCCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAA  5900
                                        insert pLM1
                                        ORF pLM1
 L  L  R  V  L  D  W  V  P  K  L  W  Y  H  L  H  T  F  L  E  K  H  S  T  S  D  F  L  I  G  P  C  F  F
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 16

TCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGG
AGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCC 6000

insert pLM1

ORF pLM1

L S C P I G I E D F R T W F I D L W N N S I I P Y L Q E G A K D G

ATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGT
TATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACA 6100

insert pLM1

ORF pLM1

I K V H G Q K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L

ACCACCTGCCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCC
TGGTGGACGGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGG 6200

insert pLM1

ORF pLM1

Y H L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S D F

TCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAA
AGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATT 6300

insert pLM1

ORF pLM1

L M A M L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L .

GGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAG
CCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTC 6400

insert pLM1

G F G N H C H P R T A E R V H Q L S . L L L S P L L F Q S T G S P

CCCCAGGAGGAGAACAGGAGGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGC
GGGGTCCTCCTCTTGTCCTCCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCG 6500

insert pLM1

A P G G E Q E G G G D E R G G T G S V C C T F E N F L G R N G V A

GTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATT
CAAACCCTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAA 6600

insert pLM1

F G N L C P L N T F T G L L . . L W G K D D S G S F P . L . V S I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page II

ACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGGCTGAACTTGCTCA 6700
TGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAGTTTTGTGACGTCGTCAAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGT

insert pLM1

T N S W A F W G G V Q K T S K H C S S S P E F S L D L T R L N L L

AAAGAAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTTCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA 6800
TTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAATGATCAAGATCTATTGACTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTT

insert pLM1

K R S R I P A H W R P L L V L D N . S . S A I P H L . R F Y L L . I

AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATA 6900
TTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTAT

T S H T S P . T . N I K . M Q L L L L T C L L Q L I M V T N K A I

GCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTA 7000
CGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAACATTCGCAAT

f1 ori

A S Q I S Q I K H F F H C I L V V V C P N S S M Y L N A . I V S V

ATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGAC 7100
TATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTG

f1 ori

N I L L K F A L N F C . I S S F F N Q . A E I G K I P Y K S K E . T

CGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT 7200
GCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTA

f1 ori

E I G L S V V P V W N K S P L L K N V D S N V K G R K T V Y Q G D

GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTT 7300
CCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAA

f1 ori

G P L R E P S P . S S F L G S R C R K A L N R N P K G S P R F R A

GACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAAC 7400
CTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTG

f1 ori

. R G K P A N V A R K E G K K A K G A G A R A L A S V A V T L R V T

CACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATA 7500
GTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTAT

f1 ori

T T P A A L N A P L Q G A S G G T F R G N V R G T P I C L F F I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

```
CATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAA
GTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAGGACTCCGCCTTTCTTGGTCGACACCTT  7600
 H  S  N  M  Y  P  L  M  R  Q  .  P  .  .  M  L  Q  .  Y  .  K  R  K  S  P  E  A  E  R  T  S  C  G

TGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCC
ACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGG  7700
 M  C  V  S  .  G  V  E  S  P  Q  A  P  Q  Q  A  E  V  C  K  A  C  I  S  I  S  Q  Q  P  G  V  E  S  P

CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGC
GGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCG  7800
 Q  A  P  Q  Q  A  E  V  C  K  A  C  I  S  I  S  Q  Q  P  .  S  R  P  .  L  R  P  S  R  P  .  L  R  .

CCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGG
GGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCC  7900
 P  V  P  P  I  L  R  P  M  A  D  .  F  F  L  F  M  Q  R  P  R  P  P  R  P  L  S  Y  S  R  S  S  E

AGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
TCCGAAAAAACCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGG  8000
                                                                                Kan/Neo
 E  A  F  L  E  A  .  A  F  A  K  I  D  Q  E  T  G  .  G  S  F  R  M  I  E  Q  D  G  L  H  A  G  S  P

GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCG
CCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGC  8100
                                          Kan/Neo
 A  A  W  V  E  R  L  F  G  Y  D  W  A  Q  Q  T  I  G  C  S  D  A  A  V  F  R  L  S  A  Q  G  R  P

GTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAG
CAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTC  8200
                                          Kan/Neo
 V  L  F  V  K  T  D  L  S  G  A  L  N  E  L  Q  D  E  A  A  R  L  S  W  L  A  T  T  G  V  P  C  A

CTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAA
GACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTT  8300
                                          Kan/Neo
 A  V  L  D  V  V  T  E  A  G  R  D  W  L  L  L  G  E  V  P  G  Q  D  L  L  S  S  H  L  A  P  A  E  K

AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGT
TCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCA  8400
                                          Kan/Neo
 V  S  I  M  A  D  A  M  R  R  L  H  T  L  D  P  A  T  C  P  F  D  H  Q  A  K  H  R  I  E  R  A  R

ACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGC
TGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCTCGTACG  8500
                                          Kan/Neo
 T  R  M  E  A  G  L  V  D  Q  D  D  L  D  E  E  H  Q  G  L  A  P  A  E  L  F  A  R  L  K  A  S  M
```

Tuesday, 18 November 1997 11:48 Page 13

fig 34 pLM4 (1 > 10070) Site and Sequence

CCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGC 8600
GGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGA

Kan/Neo

P D G E D L V V T H G D A C L P N I M V E N G R F S G F I D C G R L

GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT 8700
CCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCA

Kan/Neo

G V A D R Y Q D I A L A T R D I A E E L G G E W A D R F L V L Y G

ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC 8800
TAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGG

Kan/Neo

I A A P D S Q R I A F Y R L L D E F F . A G L W G S K . P T K R R

CAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGC 8900
GTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCG

P T C H H E I S I P P P P S M K G W A S E S F S G T P A G . S S S A

GGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAA 9000
CCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATGGCCTTCCTTGGGCGCGATACTGCCGTTATT

G I S C W S S S P T L G G G . L K H G R R Q Y R K E P A L . R Q .

AAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTG 9100
TTTCTGTCTTATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGGTGGCTCTGGGGTAAC

K D R I K R T V L G R L F I N A G F G P R A G T L S I P H R D P I

GGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCC 9200
CCCGGTTATGCGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTCGGTTGCAGCCCCGCCGTCCGGGACGG

G A N T P A F L P F P H P T P Q V R V K A Q G S Q P T S G R Q A L F

ATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCA 9300
TATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGT

. P Q V T H I Y F R L I . N F I F N L K G S R . R S F L I I S . F

AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG 9400
TTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGAC

pUC ori

K S L N V S F R S T E R Q T P . K R S K D L L E I L F F C A . S A

CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG 9500
GAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTC

pUC ori

A C K Q K N H R Y Q R W F V C R I K S Y Q L F F R R . L A S A E R F

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence Page 14

```
ATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTC 9600
pUC ori
 Y  Q  I  L  S  F  .  C  S  R  S  .  A  T  T  S  R  T  L  .  H  R  L  H  T  S  L  C  .  S  C  Y  Q

TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
ACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCAC 9700
pUC ori
 V  L  L  P  V  A  I  S  R  V  L  P  G  W  T  Q  D  D  S  Y  R  I  R  R  S  G  R  A  E  R  G  V  R

CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
GTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTG 9800
pUC ori
 A  H  S  P  A  W  S  E  R  P  T  P  N  .  D  T  Y  S  V  S  Y  E  K  A  P  R  F  P  K  G  E  R  R  T

AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC
TCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG 9900
pUC ori
 G  I  R  .  A  A  G  S  E  Q  E  S  A  R  G  S  F  Q  G  E  T  P  G  I  F  I  V  L  S  G  F  A  T

TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
AGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGAC 10000
pUC ori
 S  D  L  S  V  D  F  C  D  A  R  Q  G  G  G  A  Y  G  K  T  P  A  T  R  P  F  Y  G  S  W  P  F  A

GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 10070
 G  L  L  L  T  C  S  F  L  R  Y  P  L  I  L  W  I  T  V  L  P  P  C  I
```

SEQUENCE ID No. 8.

PAGE 1

Page 1

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

16p

```
                                                                                           Bgl I
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG  100
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC
 .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  V  P  A  V  L  T

                             Aat II                                                  Aat II
CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT  200
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA
 A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S  N  A  N  R  D  F  P  L  T  S  M  G  G  V  F  T  V

           Bgl I                Nde I                   Aat II                   Bgl I
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA  300
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT
 N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A  P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V

                                    SnaB I              Nco I
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA  400
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT
 H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H  R  Y  Y  H  G  D  A  V  L  A  V  H  Q  V  A  V

                                          Aat II
TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA  500
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT
 I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q  V  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .

                                                                                        Nhe I  Ec47
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA  600
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT
 Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G  G  L  Y  K  Q  S  V  F  S  E  P  S  D  P  L  A  L

        Nco I
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG  700
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC
                                          eGFP.C.e.unc53
 P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC  800
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG
                                          eGFP.C.e.unc53
 V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T
```

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 1

```
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG  900
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC
                                       eGFP.C.e.unc53
 T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E

                                                  KspI
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG  1000
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC
                                       eGFP.C.e.unc53
 R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA  1100
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT
                                       eGFP.C.e.unc53
 D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG  1200
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC
                                       eGFP.C.e.unc53
 V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA  1300
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT
                                       eGFP.C.e.unc53
 P  D  N  H  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G

                                 BspM II    Bgl II                                                    Asu II
                                                                                                      EcoN I
TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC  1400
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG
                                       eGFP.C.e.unc53
                                                            C.e. unc53
 I  T  L  G  M  D  E  L  Y  K  S  G  L  R  S  T  S  N  V  E  L  I  P  I  Y  T  D  W  A  N  R  H  L  S

                                                   Nru I                                              EcoR I
GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA  1500
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT
                                       eGFP.C.e.unc53
                                            C.e. unc53
 K  G  S  L  S  K  S  I  R  D  I  S  N  D  F  R  D  Y  R  L  V  S  Q  L  I  N  V  I  V  P  I  N  E
```

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 3

Bsm I

TTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT 1600
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTGTAGCTTGGACCTACCGGAGCTTTGCACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA

eGFP.C.e.unc53

C.e. unc53

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

EcoR V, Pvu II, Ear I, Ksp632I, Hind III

CGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA 1700
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCGTCTTCGAAGCCGTTGACTT

eGFP.C.e.unc53

C.e. unc53

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L K

Sst II, PmaCI, PmaCI

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCA 1800
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT

eGFP.C.e.unc53

C.e. unc53

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA 1900
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT

eGFP.C.e.unc53

C.e. unc53

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

Aat II

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAA 2000
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT

eGFP.C.e.unc53

C.e. unc53

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4

```
                EcoR V          Ear I
                                Ksp632I                                     Asu II
TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA  2100
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTTGGCTGGATGGAGGGTTGAGGTTTTT
                                        eGFP.C.e.unc53
                                         C.e. unc53
 V  G  S  T  I  S  T  S  A  K  S  L  E  S  S  S  T  Y  S  S  I  S  N  L  N  R  P  T  S  Q  L  Q  K

   Xba I                                                          Nhe I
CCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTG  2200
GGAAGATCTGGTGTTTGGGTCGATCAAGCACAACGATGTTGATGTTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGGTTTTGAAC
                                        eGFP.C.e.unc53
                                         C.e. unc53
P  S  R  P  Q  T  Q  L  V  R  V  A  T  T  T  K  I  G  S  S  K  L  A  A  P  K  A  V  S  T  P  K  L

                                                                     Bsm I
CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCATC  2300
GAAGACACTTCTGATAACCTCGTTTTGTTCTCGGGCTATTGTCGCCACCACCACCACCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG
                                        eGFP.C.e unc53
                                         C.e. unc53
A  S  V  K  T  I  G  A  K  Q  E  P  D  N  S  G  G  G  G  G  G  M  L  K  L  K  L  F  S  S  K  N  P  S

TTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG  2400
AAGGAGTAGCTTATCGGGTGTTGGATGCTCTTTCCGCCGCCGCCACGGAGTTGTTGTTTGAAACAGCTTTTAGCGACGGGGTCACTTTTCACCGGACTTC
                                        eGFP.C.e.unc53
                                         C.e. unc53
 S  S  S  N  S  P  Q  P  T  R  K  A  A  A  V  P  Q  Q  Q  T  L  S  K  I  A  A  P  V  K  S  G  L  I

               BstX I                         Hind III
CGGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAG  2500
GCCGGCTGGTCATTCGACCCTTCACGGTGCAGATACAGCTTCGAAACATGCGGTTTTCAAAGGATGGCATTTTGCCTGCGGGGTTAGTATAGAGTTGTTC
                                        eGFP.C.e.unc53
                                         C.e. unc53
P  P  T  S  K  L  G  S  A  T  S  M  S  K  L  C  T  P  K  V  S  Y  R  K  T  D  A  P  I  :  S  Q  Q
```

Ear I
Ksp632I BspM II

```
ACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT 2600
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTTCTCAGGCCTATGCGACCTAAGTTGTCGTGCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTCGTA
```

eGFP.C.e.unc53

C.e. unc53

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

Bsm I
Sph I
Ava III
Nsi I

```
GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCG 2700
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTCGATAGTCTGTCGGC
```

eGFP.C.e.unc53

C.e. unc53

H S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

Ssp I

```
ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC 2800
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTCGGACAACTCCTTTTTGGTTGTGACCGTCACTTTCCTCACTTTTCGTGTCGCTTTTTTCTAG
```

eGFP.C.e.unc53

C.e. unc53

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

PmaCI
Pvu II PmaCI EcoR V

```
CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA 2900
GTGGAGGTCGACAAGGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGTATTCTTCAACTGTTTACTGGGGCACTATAGACT
```

eGFP.C.e.unc53

C.e. unc53

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

AlwN I

```
AAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA 3000
TTTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAGT
```

eGFP.C.e.unc53

C.e. unc53

K P E P E K L C S M S I D T T D V P P L P P L K S V V P L K M T S

Tuesday, 18 November 1997 10:34 Page 6
fig 30 pEGFP72 (1 > 9697) Site and Sequence

SplI
SplI

ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACT 3100
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTTAGTGTAGCGGACAGTTCAGCAAACCTATACTCGTCAGCAGGCGCAGACTTCTGA

eGFP.C.e.unc53

C.e. unc53

I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC 3200
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTTTGAAGACCATTAGTAAGCGACCTCTCTTCCTACCCTTTCTTATTCTGTAGTCTTAG

eGFP.C.e.unc53

C.e. unc53

S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S

CAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT 3300
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTTACTCCCTCTTCGACTTTCTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA

eGFP.C.e.unc53

C.e. unc53

S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A Q N G Y

Asu II Sst I BspM II

CCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGG 3400
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCCTGCTAAACAGGCCTCATCTGTACC

eGFP.C.e.unc53

C.e. unc53

P D N F E D S S S L S S G I S D N N E L D D I S T D D L S G V D M

CAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTCGGTCCTCCACATCAGT 3500
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAAACAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA

eGFP.C.e.unc53

C.e. unc53

A T V A S K H S D Y S H F V R H P T S S S S K P R V P S R S S T S V

Xho I; Bgl I; Nar I; Bbe I

CGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGGACAA 3600
GCTAAGAGCTAGAGCTCGTCTTGTCCTCTTACACATGTTTGAAGACAGGGTCACGGCTTGCTCGGTTGCACCGCGGCGACGGTGGAGTTGGAAGCCTGTT

eGFP.C.e.unc53

C.e. unc53

D S R S R A E Q E N V Y K L L S Q C R T S Q R G A A A T S T F G Q

Xma I; Sma I; Pvu II; Spe I; Sal I

CATTCGCTAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTATGCACTCACAGACTAGTCGACGAC 3700
GTAAGCGATTCTAGGGGCCCTATGAGTAGGATAAGAGGTGTGAATAGTCACAGTCGACTATTCCTGTGTTACAGATACGTGAGTGTCTGATCAGCTGCTG

eGFP.C.e.unc53

C.e. unc53

H S L R S P G Y S S S Y S P H L S V S A D K D T M S M H S Q T S R R

CTTCTTCACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATCCACCGAGCACAGAATGGCGGCTCT 3800
GAAGAAGTGTTTTTGGTTCGATAAGTCCGGTTAAAGTAAGTGAACTAGCATTTACGGTGGAAGTTCTCAAGTGTAGGTGGCTCGTGTCTTACCGCCGAGA

eGFP.C.e.unc53

C.e. unc53

P S S Q K P S Y S G Q F H S L D R K C H L Q E F T S T E H R M A A L

Bam HI

CTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCCTACTCGGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGA 3900
GAACTCGGGCTCTGCCCACGGCTTGAGCTACAGCTTTATACTAAGAAGTCCTAGGATGAGCCGCGCAAGGGCTCCACCTTCGAGATGACCATAGATACCT

eGFP.C.e.unc53

C.e. unc53

L S P R R V P N S M S K Y D S S G S Y S A R S R G G S S T G I Y G

Bam HI; Nhe I; Nde I

GAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGCTAGCACGACAGCAT 4000
CTCTGCAAGGTTGACGTGTCTGATAGGCTACTTTTTAGGGGGCGTGTAAGACGGTTTTCACTCTACCCTAGGGTTGATAGTGACCGATCGTGCTGTCGTA

eGFP.C.e unc53

C.e. unc53

E T F Q L H R L S D E K S P A H S A K S E M G S Q L S L A S T T A

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 8

```
                                                                PmaCI
                                                                   PmaCI                          Sal I
ATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGTCGACTCACTAACCAAGAAACAGGA
TACCTAGAGAGTTACTCTTCATGCTTGTACGATAAGCCCTGTACCGTGCACTGAACCTCACAATGTTCTTGTGACAGCTGAGTGATTGGTTCTTTGTCCT  4100
                                   eGFP.C.e.unc53
                                   C.e. unc53
 Y  G  S  L  N  E  K  Y  E  H  A  I  R  D  M  A  R  D  L  E  C  Y  K  N  T  V  D  S  L  T  K  K  Q  E

                                                                                         Ear I
                                               Hind III                   Cla I          Ksp632I
GAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACGATTC
CTTGATACCTCGTAACAAACTAGAAAAACTCGTTTTCGAATCTTTTGAGTGAGTTGTGTAACTAGCTAGGTTGAACTTCGGACTTCTCCGTTATGCTAAG  4200
                                   eGFP.C.e.unc53
                                   C.e. unc53
  N  Y  G  A  L  F  D  L  F  E  Q  K  L  R  K  L  T  Q  H  I  D  R  S  N  L  K  P  E  E  A  I  R  F

AGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCTTCGTCAACCATCTC
TCCGTCCTGTAACGAGTAAACTCCCTATAATCGTTAGTAGAACGTAGGTTGAGTCGAGTACGATTGCTTCCGCGACCACTCGAAGAAGCAGTTGGTAGAG  4300
                                   eGFP.C.e.unc53
                                   C.e. unc53
 R  Q  D  I  A  H  L  R  D  I  S  N  H  L  A  S  N  S  A  H  A  N  E  G  A  G  E  L  L  R  Q  P  S

                                                                                            Ear I
                     Cla I   Cla I                                                  Sst I   Ksp632I
TGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGCTCGTTTGGCAAGAACAAGAAGAG
ACCTTAGTCAACGTAGGGTAGCTAGTAGCTACAGTAGCAGCAGCTTTTCGTCGTTCGTCCTCTTCTAGTCGAACTCGAGCAAACCGTTCTTGTTCTTCTC  4400
                                   eGFP.C.e.unc53
                                   C.e. unc53
L  E  S  V  A  S  H  R  S  S  M  S  S  S  S  K  S  S  K  Q  E  K  I  S  L  S  S  F  G  K  N  K  K  S

  Bam HI                                                                 Nde I       BspM II
CTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAACTCTTGAC
GACCTAGGCGAGGAGTGAGAGGTTCAAGTGGTTCTTCTTCTTGTTCTTGATGCTGCTTCGTGTATACGGTAGTTAAAGGCCTAGAGTTCCTTGAGAACTG  4500
                                   eGFP.C e unc53
                                   C.e. unc53
 W  I  R  S  S  L  S  K  F  T  K  K  K  N  K  N  Y  D  E  A  H  M  P  S  I  S  G  S  Q  G  T  L  D
```

Sst I ApaL I

AACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTTGATG 4600
TTGTAACTACACTAACTCAACTTCGTTCTCGAGTTTCTTGCGCTATCACGTGAAATGCTTCAGGCGGAACTGTTAGACCTAGCACGGGCGCTTCAACTAC

eGFP.C.e.unc53

C.e. unc53

N I D V I E L K Q E L K E R D S A L Y E V R L D N L D R A R E V D

TTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCG 4700
AAGACTCCCTCTGTCACTTGTTCAACTTTTGGCTCTTGTTCGTTAATTTCTTTCTTCACCTGTTTGAGTGGTTGCCAGGTCGGTGAGCACGAAGAAGGGC

eGFP.C.e.unc53

C.e. unc53

V L R E T V N K L K T E N K Q L K K E V D K L T N G P A T R A S S R

KspI BsrI Asu II

CGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAAC 4800
GCGGAGTTAAGGTCAATAGATGCTGCTACTCGTACAGATACTACGTCGCACATCGTCATGTAGTCGATCAGTTAGAAGCTTTGCTAGGAGACCGACGTTG

eGFP.C.e unc53

C.e. unc53

A S I P V I Y D D E H V Y D A A C S S T S A S Q S S K R S S G C N

Pvu I Hpa I EcoR V

TCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCA 4900
AGTTAGTTCCAATGACATTTGCACCTGTAGCGACCTCTTTAGTCAAGCTAGCAATTGGGCCTGTTTCTCTATTAGCATCCTATAGAACGGTACAGTTGGT

eGFP.C.e.unc53

C.e. unc53

S I K V T V N V D I A G E I S S I V N P D K E I I V G Y L A M S T

Cla I

GTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATCAACTTGGAATCGATGCTCC 5000
CAGTCAGTACGACCTTTCTGTAACTACAAAGATAAGATCCTGATAAACTTCAGATGGATAGGTCTTAACTACACCTCGTAGTTGAACCTTAGCTACGAGG

eGFP.C e unc53

C.e. unc53

S Q S C W K D I D V S I L G L F E V Y L S R I D V E H Q L G I D A P

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 10

Mlu I

TGATTCTATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCATCCAACTGACATTCTTACTTCCTCA 5100
ACTAAGATAGGAACCGATAGTTTAACCACTTGAAGCTGCGCAGTAACCTCTGAGGTGTTGGTACTATTGGTCGGTAGGTTGACTGTAAGAATGAAGGAG

eGFP.C e.unc53

C.e. unc53

D S I L G Y Q I G E L R R V I G D S T T M I T S H P T D I L T S S

ACTACAATCCGAATGTTCATGCACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACTCGTCA 5200
TGATGTTAGGCTTACAAGTACGTGCCACGGCGTGTCTCAGCGCATCTGTCAGACCAGGAACTATACGAAGAAGGTTTCGTTTACTAAGAGGTTGAGCAG

eGFP.C.e.unc53

C.e. unc53

T T I R M F M H G A A Q S R V D S L V L D M L L P K Q M I L Q L V

Aat II Bsrl Bsrl Asu II

AGTCAATTTTGACAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTATGTATCTATTCGAAC 5300
TCAGTTAAAACTGTCTCTCTGCAGACCACAATCGACCTCGTTGACCTTAACCTTTCTCGTTTGACCGCTTCTGGGACCGACGAATACATAGATAAGCTTG

eGFP.C.e.unc53

C.e. unc53

K S I L T E R R L V L A G A T G I G K S K L A K T L A A Y V S I R T

Bsm I Xmn I Bgl II

AAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGC 5400
TTTAGTTAGGCTTCTATCATAACAATTATAGTCGTAAGGACTTTTGTTATTTCTTCTTAACGAAGTTCACCTTGCTGCGGACCTTTTCTAGAACTCTTCG

eGFP.C.e.unc53

C.e. unc53

N Q S E D S I V N I S I P E N N K E E L L Q V E R R L E K I L R S

Ava III
Nsi I Xba I

AAAGAATCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGTCCCACTTCAAAACAACGAAGGTC 5500
TTTCTTAGTACGTAGCATTAAGATCTATTATAGGGTTTCTTAGCTTAACGTAAACAACATAGGCAAAAACGTTTACAGGGTGAAGTTTTGTTGCTTCCAG

eGFP.C.e.unc53

C.e. unc53

K E S C I V I L D N I P K N R I A F V V S V F A N V P L Q N N E G

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence
Page 11

EcoR V

CATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAAATGTCAGTAATGTCGAATCGTCTCGAAGGATTCAT
GTAAACATCATACGTGTCAGTTGGCTATAGTTTAGGGACTCGAAGTTTAAGTGGTGTTAAAGTTTTACAGTCATTACAGCTTAGCAGAGCTTCCTAAGTA 5600

eGFP.C.e.unc53

C.e. unc53

P F V V C T V N R Y Q I P E L Q I H H N F K M S V M S N R L E G F I

Sst I
Ear I
Ksp632I

CCTACGTTACCTCCGACGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCT
GGATGCAATGGAGGCTGCTGCCCGCCATCTCCTACTCATAGCAGATTGACATGTCTACGGTAGTCTCGAGAAGTTTTAGTAACTGAAGAAGGGTTATCGA 5700

eGFP.C.e.unc53

C.e. unc53

L R Y L R R R A V E D E Y R L T V Q M P S E L F K I I D F F P I A

Ear I
Ksp632I
EcoR I
Sph I
Bam HI

CTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACTGTCCTCTAACTGTCGATGGATCCC
GAAGTCCGGCAGTTATTAAAATAACTCTTTTGCTTAAGACAACTACACTGTCAACCAGGTTCTCGTACGAACTTGACAGGAGATTGACAGCTACCTAGGG 5800

eGFP.C.e.unc53

C.e. unc53

L Q A V N N F I E K T N S V D V T V G P R A C L N C P L T V D G S

GTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCACTTCCTTCGA
CACTTACCAAGTAAGCTAACACCTTACTCTTGAAGTAAGGTATAAACCTTGCACAACGATCTCTACCGTTTTTTTGGAAGCCAGCGACGTGAAGGAAGCT 5900

eGFP.C.e.unc53

C.e. unc53

R E W F I R L W N E N F I P Y L E R V A R D G K K T F G R C T S F E

Bam HI
Tth I
Tth I

GGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAAGACCTCGTCCCGTCA
CCTAGGGTGGCTGTAGCAGAGATTTTTTACCGGCACCAAGCTACCACTTTTGGGCCTCTTACACGAGTTTGCAGAAGTTGAGGTTCTGGAGCAGGGCAGT 6000

eGFP.C.e.unc53

C.e. unc53

C P T D I V S K K W P W F D G E N P E N V L K R L Q L Q D L V P S

Tuesday, 18 November 1997 10:35 Page 17

fig 30 pEGFP72 (1 > 9697) Site and Sequence

BspM I Xho I Sph I

CCTGCCAACTCATCCCGACAACACTTCAATCCCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACT 6100

GGACGGTTGAGTAGGGCTGTTGTGAAGTTAGGGGAGCTCAGCAACTAGGTTAACGTACGATGGTTCGTAGTCTGGTAGCTGTTGTAAACTTGTCTTCTGA

eGFP.C.e.unc53

C.e. unc53

P A N S S R Q H F N P L E S L I Q L H A T K H Q T I D N I . T E D

Asp 718 Kpn I

CTAATCTTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCCCCCTTTTCCCCCCAATTTCCCAGAACCTCCT 6200

GATTAGAAGAGAGCGGAGAGGGGGCGAAAGGAATAGAAGCATGGCCATGGACTACTAAGGGGTAAAAGGGGGAAAAGGGGGGTTAAAGGGTCTTGGAGGA

S N L L S P L P R F P Y L R T G T . . F P I F P L F P P I S Q N L L

Xma I Sma I Dra I Xmn I

GTTCCCTTTGTTCCTAGTCCTCCCGGGTGCCGACGCCGAAGCGATTTAAAAACCTTTTTCTTTCCGAAACATTTCCCATTGCTCATTAATAGTCAAATTG 6300

CAAGGGAAACAAGGATCAGGAGGGCCCACGGCTGCGGCTTCGCTAAATTTTTGGAAAAAGAAAGGCTTTGTAAAGGGTAACGAGTAATTATCAGTTTAAC

F P L F L V L P G A D A E A I . K P F S F R N I S H C S L I V K L

Xma I Apa I Sma I Xho I Bam HI Xba I Bcl I

AATAAACAGTGTATGTACTTAAAAAAAAAAAAAAAAAAAACTCGAGGGGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACA 6400

TTATTTGTCACATACATGAATTTTTTTTTTTTTTTTTTTTGAGCTCCCCCCGGGCCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGGTGT

N K Q C M Y L K K K K K K L E G G P G I H R I . I T D H N Q P Y H

Dra I Bsm I Hpa I

TTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAG 6500

AAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTC

I C R G F T C F K K P P T P P P E P E T . N E C N C C C . L V Y C S

Bsm I

CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGT 6600

GAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACA

L . V L Q I K Q . H H K F H K . S I F F T A F . L W F V Q T H Q C

Mlu I Ssp I

ATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAA 6700

TAGAATTGCGCATTTAACATTCGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTT

I L T R K L . A L I F C . N S R . I F V K S A H F L T N R P K S A

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

BsrI

AATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAA
TTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTT 6800
K S L I N Q K N R P R . G . V L F Q F G T R V H Y . R T W T P T S

Dra III

GGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACC
CCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGG 6900
G E K P S I R A M A H Y V N H H P N Q V F W G R G A V K H . I G T

Nae I

CTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGC
GATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCG 7000
L K G A P D L E L D G E S R R T W R E R K G R K R K E R A L G R W

KspI KspI

AAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
TTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCT 7100
Q V . R S R C A . P P H P P R L M R R Y R A R Q V A L F G E M C A E

BspH I Ssp I Ear I Ksp632I

ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTC
TGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAG 7200
P L F V Y F S K Y I Q I C I R S . D N N P D K C F N N I E K G R V

OxaN I Pvu II Sph I Ava III Nsi I

CTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAAT
GACTCCGCCTTTCTTGGTCGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTA 7300
L R R K E P A V E C V S V R V W K V P R L P S R Q K Y A K H A S Q

Sph I Ava III Nsi I

TAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAA
ATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATT 7400
L V S N Q V W K V P R L P S R Q K Y A K H A S Q L V S N H S P A P N

BsrI Nco I Bgl I Sfi I

CTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
GAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCG 7500
S A H P A P N S A Q F R P F S A P V L T N F F Y L C R G R G R L G

Tuesday, 18 November 1997 10:35 Page 1
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Stu I
Avr II
Cla I

CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGA 7600
GAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACT
L . A I P E V V R R L F W R P R L L Q R S I K R Q D E D R F A . L

BspM I
Xma III

ACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTG 7700
TGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCAC
N K M D C T Q V L R P L G W R G Y S A M T G H N R Q S A A L M P P C

Nar I
Bbe I
KspI

TTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGC 7800
AAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCG
S G C Q R R G A R F F L S R P T C P V P . M N C K T R Q R G Y R G

Bal I
Fsp I
Pvu II
Tth I

TGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCT 7900
ACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGA
V P R R A F L A Q L C S T L S L K R E G T G C Y W A K C R G R I S

BspM I

GTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAA 8000
CAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTT
C H L T L L L P R K Y P S W L M Q C G G C I R L I R L P A H S T T I

Ear I
Ksp632I

GCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAC 8100
CGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTG
R N I A S S E H V L G W K P V L S I R M I W T K S I R G S R Q P N

Sph I
Nco I

TGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTT 8200
ACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAA
C S P G S R R A C P T A R I S S . P M A M P A C R I S W W K M A A

Nae I
Rsr II
Ear I
Ksp632I

TTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGG 8300
AAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACC
F L D S S T V A G W V . R T A I R T . R W L P V I L L K S L A A N G

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGG
CGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCC 8400
L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S G

Asu II BspM I
GTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGG
CAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCC 8500
V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F P

Nae I Kspl Avr II
GACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGG
CTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATGGCC 8600
G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T G

Kspl
AAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTC
TTCCTTGGGCGCGATACTGCCGTTATTTTTCTGTCTTATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAG 8700
R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A L

TGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCA
ACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGT 8800
C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A R

Alwn I OxaN I Dra I Dra I
GCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGA
CGGTTGCAGCCCCGCCGTCCGGGACGGTATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACT 8900
S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G E

BspH I
AGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA
TCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCT 9000
D P F . . S H D Q N P L T . V F V P L S V R P R R K D Q R I F L R

TCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCG
AGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGC 9100
S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F S

Bsrl
AAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT
TTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTA 9200
E G N W L Q Q S A D T K Y C P S S V A V V R P P L Q E L C S T A Y I

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 16

Alwn I

ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA 9300
TGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGT
P R S A N P V T S G C C Q V R . V V S Y R V G L K T I V T G . G A

ApaL I

GCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC 9400
CGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGG
A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K R

ACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC 9500
TGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAG
H A S R R E K G G Q V S G K R Q G R N R R A H E G A S R G K R L V S

TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGC 9600
AAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCG
L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R G

Ava III
Nsi I

CTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 9697
GAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Monday, 1 December 1997 14:12
fig pCB501

SEQUENCE I.D. No 9

Page 1

10 20 30 40 50 60 70

ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCT 70

CGAGGATCAGAAGAAATTGGACCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAA 140

TTTTGAGTTTACGACTACAAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCTATTCTGTATGACT 210

AGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAAAGGAACATTATTTACTTTGCTTATTTGCCCTAAC 280

TTTGATTTAGTTTTTCGATCAACTAGATCTTACAAAACTTGCAATACAATTCCATTTTCAGATTACCCTC 350

360 370 380 390 400 410 420

GCCACGTGTCGCCACGTCAGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACA 420

TCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGT 490

ATAAAATGCCCAGAAAAGAAGATGATAAAAATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTCTAA 560

TATGACGGCTTATATCTCAATTTTCTTGAGTTTTATCAAAAAATTTTCCACTATACAAATGTAGAAAAGT 630

ATTTTGCACAAATTTTGTCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAA 700

710 720 730 740 750 760 770

AGTGGAAGGAGCGCAAGTCTATACTGGAAATAATGATCTGAAACAAATTTGTGCTATTCTCAAATGTTTA 770

AGACATGTTTTGAAGATTTTTTCAAATTCGCACTAGTTTCAGAACCTTCCTTTTTGTATGAAAAAGTAAA 840

AAAAAACTATTTCAAACCCTCACCGCCACCATGTTTCAACTCTTAATTTTTATAAAATTTTGCAATTTAC 910

AAATCGCCTCCCCTTGCCCGAAAAGTGCCCACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATT 980

CATGTGAGAAAACACAGAAGAGGCTAACTAAATTGACAGGGACAGGTTGTCCCTCTTCTCCCTCCTTCTC 1050

1060 1070 1080 1090 1100 1110 1120

CCGCCTCCTCCTCCCGTTCCATCTCCAACAACAACAATTTTCCAATTTCGTTGTCCATTTTGCTTATAAA 1120

CATTTGTGTGTGGAAGGAAACTACACGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACTC 1190

TTTCGGAAATTGATGAATAAAGATACACCCGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTT 1260

TTTCATCGTCTCAACTTGCGCATGAGTCTTCCCCCGCTCTCATCACTGACAATTAATGTCGGGTTTTATG 1330

CGCTCTTTCCTATTCCGCCACTCATTCTGGGTTACCACAAACTGGAATACATTTTACTACTATTCAAGCC 1400

1410 1420 1430 1440 1450 1460 1470

ATTTATTTTGATATTTAATTTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACG 1470

ATATTTTCGATTTTAAAAAATCTGAAAAGTTTCAAAAAATCAATAAATATTCCCTAACAAATTGTATGCC 1540

TAAAATTTTATTTCTACTGTTGACAATATCTTTATATGTATCACTGTTTTCCATCTCAAAACCTTGAATC 1610

CCCCAAGTTATAGGAACCTCCGTGTCACATTTCCCATGCTATGAATCGCTACTCAGCACATATCCAAAAA 1680

TTAAGCTAGACGGTTGATAATTATTGGGCACGCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAAGC 1750

1760 1770 1780 1790 1800 1810 1820

ACAGATTATCTATCAAATTCAATCTTTTGAACATTCACCCAGTTCGTACAATTTTCCATGCTTTTTGGCCC 1820

ATTAAAAAACTTTCTCACCTCTTCATCCATCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACTCT 1890

ACTTTTTAAGAGAAGGAGATACTGAGCCACATGGCGTGTGACCCTTTTCATCTCGTCCGTTCGGTCTCAA 1960

ATTCACGCTCATACTAACTCTTCAAATAGCCATAGACCTCCTTGTTTTCTTCTTCGTTTTGACTCGCGCC 2030

TATTTTTTGTGGCTGCCTGAAAGCCGGGAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100

2110 2120 2130 2140 2150 2160 2170

AAAAACGAGGCAATTTAAAAATATTAAAATTAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAA 2170

ATAAAACAAATAGGAACCGCCAGATCAAAATTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGG 2240

CTGAACAGAAAACTGAAGTGCCTGCATAAATCTAGTGTAACGTTTAGATTGAACTCGGAAATCCTAAGCC 2310

TGAACTATAGCCTTATTCTAGATCTTAGTTGCGCATAAGCTCAAGCCCAAGCAGAAATGACTTCCATTTA 2380

GTTAAGCCTAGATTGACTTGCTTGCTTCAGTCTAATCCAGACTAGATTTCCAAGAGACTTTTCAATTTT 2450

Page 2

Monday, 1 December 1997 14:12

Fig pCB501

```
      2460      2470      2480      2490      2500      2510      2520
AAATGTTTCCAGTTTCTTGTTACTTAAAATCTTAATGCCCTGTGATGCGTAAAATCGTTATCCCTTTCTC 2520
TCACACTTTCAATTACAGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTC 2590
ATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCCTTCGAGTGGCAATAATAAT 2660
GTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTATCCGATCCTTCCGGCTTCTTTTTAGAAATT 2730
ATATTATTTCAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800
      2810      2820      2830      2840      2850      2860      2870
AAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTA 2870
GCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTGCTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCG 2940
ATAACAGCGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCATCTTCCTC 3010
ATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCT 3080
GCCCCAGTGAAAAGTGGCCTGAAGCCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTT 3150
      3160      3170      3180      3190      3200      3210      3220
GTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAGACTCGAAACGATGCTC 3220
AAAGAGCAGTGAAGAAGAGTCCGGATACGCTGCATTCAACAGCACGTCGCCAACGTCATCATCGACGGAA 3290
GCTTCCCTAAGCATGCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACG 3360
ATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAA 3430
TATTATCAACAAGCCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAA 3500
      3510      3520      3530      3540      3550      3560      3570
GATCCACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCAC 3570
ATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACCAGAACCTGAAAAGCTCCAATCAATGAGCAT 3640
TGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCAATCCGA 3710
CAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATG 3780
AGCAGTCGTCCGCGTCTGAAGACTCCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAAC 3850
      3860      3870      3880      3890      3900      3910      3920
TTCTGGTAATCATTCGCTCGAGAGAAGGATGGGAAAGAATAAGACATCAGAATCCAGCGGCTACACCTCT 3920
GACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAG 3990
CACAGAACGGCTATCCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTCGGAATATCCGATAACAACGA 4060
GCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGGCAACAGTCGCCTCCAAACATAGCGAC 4130
TATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTCGGTCCTCCACAT 4200
      4210      4220      4230      4240      4250      4260      4270
CAGTCCATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCA 4270
ACGTGGCGCCGCTGCCACCTCAACCTTCGGACAACATTCGCTAAGATCCCCGGGATACTCATCCTATTCT 4340
GCACACTTATCAGTCTCAGCTGATAAGGACACAATGTCTATCCACTCACAGACTAGTCGACGACCTTCTT 4410
CACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATC 4480
CACCGACCACAGAATGGCGGCTCTCTTGACCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCT 4550
      4560      4570      4580      4590      4600      4610      4620
TCAGGATCCTACTCGGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGAGAGACGTTCCAACTGC 4620
ACAGACTATCCGATGAAAAATCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGC 4690
TAGCACGACAGCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGCACATGGCACGTCACTTG 4760
GAGTGTTACAAGAACACTGTCGACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTGTTTGATCTTT 4830
TTGAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACG 4900
```

Monday, 1 December 1997 14:12

flg pCB601

4910 4920 4930 4940 4950 4960 4970

ATTCAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAAC 4970
GAAGGCGCTGGTGAGCTTCTTCGTCAACCATCTCTGCAATCAGTTGCATCCCATCGATCATCGATGTCAT 5040
CGTCGTCGAAAAGCAGCAAGCAGGACAAGATCAGCTTGAGCTCGTTTGGCAAGAACAAGAAGAGCTGGAT 5110
CCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATT 5180
TCCGGATCTCAAGGAACTCTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATA 5250

5260 5270 5280 5290 5300 5310 5320

GTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTTGATGTTCTGAGGGAGACAGT 5320
GAACAAGTTGAAAACGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACT 5390
CGTGCTTCTTCCCGCGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCA 5460
GTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAACTCAATCAACGTTACTGTAAACGTGGA 5530
CATCGCTGGAGAAATCAGTTCGATCGTTAACGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGC 5600

5610 5620 5630 5640 5650 5660 5670

AAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTT 5670
CTAAAATGGACCCAGCCTCTACCCTCGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGT 5740
GAAACGAGTGTTGGATGCAGACCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTC 5810
TCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGCTGTTCGAGACGCTGATCCCCAAGCCGATGA 5880
TGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAA 5950

5960 5970 5980 5990 6000 6010 6020

GACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATC 6020
GTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGCATCTGCAACTGTATCTTTCCAACCTAGCCAACC 6090
AGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGG 6160
CTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACC 6230
ACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCA 6300

6310 6320 6330 6340 6350 6360 6370

ACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGA 6370
CATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCAC 6440
AGGTTCCTTGAG...GCAGCAGC...TTCCTCATCGCCCCTTGGTTCT...CTGCTGGGATTG 6510
GCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGG 6580
AGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGAC 6650

6660 6670 6680 6690 6700 6710 6720

ACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAACCTGTACCACCTGCCCCCACCCACCGTGGCCC 6720
CTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGA 6790
TCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACC 6860
ATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACACCAGAAC 6930
GCTGCCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGC 7000

7010 7020 7030 7040 7050 7060 7070

ACGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTC 7070
CTAGGAACGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAAT 7140
GACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTT 7210
CTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGG 7280
CTGAACTTGCTCAAAAGAAGCCGAATTCCAGCACACTGGCCTCCCCATGGTATTCATATC GAGCTCCGC 7350

Monday, 1 December 1997 14:12 Page 4

Ilg pCB501

7360 7370 7380 7390 7400 7410 7420

ATCGGCCGCTGTCATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAAC 7420
ATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTTCTTA 7490
ATTTCTTTGTTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATT 7560
AATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCT 7630
ACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAA 7700

7710 7720 7730 7740 7750 7760 7770

AAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCA 7770
CGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTC 7840
AATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCA 7910
AGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGG 7980
AAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCT 8050

8060 8070 8080 8090 8100 8110 8120

ATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGG 8120
CCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGG 8190
AGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGT 8260
TGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGG 8330
TGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGC 8400

8410 8420 8430 8440 8450 8460 8470

CTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATG 8470
TGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACC 8540
CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATT 8610
CCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTG 8680
AAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTT 8750

8760 8770 8780 8790 8800 8810 8820

TCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT 8820
ATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC 8890
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAG 8960
TGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC 9030
AACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACG 9100

9110 9120 9130 9140 9150 9160 9170

AGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTAC 9170
TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCG 9240
GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG 9310
CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTAT 9380
GGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA 9450

9460 9470 9480 9490 9500 9510 9520

GTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCC 9520
TTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGA 9590
AAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCA 9660
CCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCA 9730
GCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGT 9800

Monday, 1 December 1997 14:12

fig pCB501

Page 5

```
         9810      9820      9830      9840      9850      9860      9870
AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGT  9870
CTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGT  9940
GCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAG  10010
CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGC  10080
ACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTG  10150
        10160     10170     10180     10190     10200     10210     10220
AGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT  10220
ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT  10290
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAG  10360
TGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATG  10430
CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC  10500
        10510     10520     10530     10540     10650     10560     10570
ACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGAT  10570
AACAATTTCACACAGGAAACAGCT  10594
```

SEQUENCE ID NO 10

fig 13 6 page

Page

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence
Enzymes : 100 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

```
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT 100
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
 T  D  R  E  I  S  R  S  P  M  V  D  S  Q  Y  N  L  L  .  C  R  I  V  K  P  V  S  A  P  C  L  C  V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG 200
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACTGGCTGTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGC
 G  G  R  .  V  V  R  E  Q  N  L  S  Y  N  K  A  R  L  D  R  Q  L  H  E  E  S  A  .  G  .  A  F  C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA 300
GACGAAGCGCTACATGCCCGGTCTATATGCGCAACTGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT
 A  A  S  R  C  T  G  Q  I  Y  A  L  T  L  I  I  D  .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT 400
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCA
 G  V  P  R  Y  I  T  Y  G  K  V  P  A  W  L  T  A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC 500
TTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG
 N  A  N  R  D  F  P  L  T  S  M  G  G  L  F  T  V  N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A

CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA 600
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT
 P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V  H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA 700
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTT
 R  Y  Y  H  G  D  A  V  L  A  V  H  Q  W  A  W  I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q

TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG 800
ACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTC
 W  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .  Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC 900
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG
                                                                  T7 promotor priming site
 G  L  Y  K  Q  S  S  L  A  N  .  R  T  H  C  L  L  A  Y  R  N  .  Y  D  S  L  .  G  D  P  S  W  L  A

GTTTAAACTTAAGCTTACCATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGAC 1000
CAAATTTGAATTCGAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG
                                    ProBond binding domain
 F  K  L  K  L  T  M  G  G  S  H  H  H  H  H  H  G  M  A  S  M  T  G  G  Q  Q  M  G  R  D  L  Y  D
```

Tuesday, 18 November 1997 10:09 Page

fig 13 pCB201 (1 > 5082) Site and Sequence

GATGACGATAAGGTACCTAGGATCCATATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACA 100

CTACTGCTATTCCATGGATCCTAGGTATACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGT

pCB201 insert = U4

U4 ORF

D D D K V P R I H M P P C R R G V N N I S V S L K G L K E K C V D

GCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCAC 200

CGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTG

pCB201 insert = U4

U4 ORF

S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S G P S G T

GGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAG 300

CCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTC

pCB201 insert = U4

U4 ORF

G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T F N M H Q

CAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATG 400

GTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTAC

pCB201 insert = U4

U4 ORF

D S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L V I L L D

ACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGT 500

TGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACA

pCB201 insert = U4

U4 ORF

D _ S E A G S I S E L V N G A L T C K Y H K C P Y I I G T T N Q P V

AAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGG 600

TTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCC

pCB201 insert = U4

U4 ORF

K M T P N H G L H L S F R M L T F S N N V E P A N G F L V R Y L F

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

```
AGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCT
TCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGA 1700
                                  pCB201 insert = U4
                                       U4 ORF
 R  K  L  V  E  S  D  S  D  I  N  A  N  K  E  E  L  L  R  V  L  D  W  V  P  K  L  W  Y  H  L  H  T

TCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCT
AGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGA 1800
                                  pCB201 insert = U4
                                       U4 ORF
 F  L  E  K  H  S  T  S  D  F  L  I  G  P  C  F  F  L  S  C  P  I  G  I  E  D  F  R  T  W  F  I  D  L

GTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTC
CACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAG 1900
                                  pCB201 insert = U4
                                       U4 ORF
 W  N  N  S  I  I  P  Y  L  Q  E  G  A  K  D  G  I  K  V  H  G  Q  K  A  A  W  E  D  P  V  E  W  V

CGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCC
GCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGG 2000
                                  pCB201 insert = U4
                                       U4 ORF
 R  D  T  L  P  W  P  S  A  Q  Q  D  Q  S  K  L  Y  H  L  P  P  P  T  V  G  P  H  S  I  A  S  P  P

AGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTC
TCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAG 2100
                                  pCB201 insert = U4
                                       U4 ORF
 E  D  R  T  V  K  D  S  T  P  S  S  L  D  S  D  P  L  M  A  M  L  L  K  L  Q  E  A  A  N  Y  I  E  S

TCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTA
AGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGAT 2200
                                  pCB201 insert = U4
                     U4 ORF
 P  D  R  E  T  I  L  C  P  N  L  Q  A  T  L  .  G  F  G  N  H  C  H  P  R  T  A  E  R  W  H  Q  L
```

Tuesday, 18 November 1997 10:09 Page
fig 13 pCB201 (1 > 5082) Site and Sequence

TCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTT 2300
AGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAA

pCB201 insert = U4

S . L L L S P L L F Q S T G S P A P G G E Q E G G G D E R G G T G

CTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACT 2400
GAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGA

pCB201 insert = U4

S V C C T F E N F L G R N G G V A F G N L C P L N T F T G L L . . L

TTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGC 2500
AACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAGTTTTGTGACG

pCB201 insert = U4

V G K D D S G S F P . L L V S I T N S V A F V G G V Q K T S K H C

AGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGGAGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCC 2600
TCGTCAAGGATTTACTAAGAGTGTTCGTTGGGACTCTCTCTGTCAGAACACTCCCTCTAGACCCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGTCTGGG

pCB201 insert = U4

S S S . M I L T S N P E R D S L V R E I V G R Q E A P Q I F S Q T

TTCCCAATTCCATCACCACTGCCAACACTCGTCCGGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTG 2700
AAGGGTTAAGGTAGTGGTGACGGTTGTGAGCAGGCCTTAAGACGTCTATAGGTCGTGTCACCGCCGGCGAGCTCAGATCTCCCGGGCAAATTTGGGCGAC

pCB201 insert = U4

L P N S I T T A N T R P E F C R Y P A Q V R P L E S R G P V . T R .

ATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC 2800
TAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCACGGAAGGAACTGGGACCTTCCACGGTGAGGGTGACAGGAAAGG

S A S T V P S S C Q P S V V C P S P V P S L T L E G A T P T V L S

TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA 2900
ATTATTTTACTCCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCCACCCCACCCCGTCCTGTCGTTCCCCCTCCTAACCCTTCTGT

. . N E E I A S H C L S R C H S I L G G G V G Q D S K G E D V E D

ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGC 3000
TATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCCTTTCTTGGTCGACCCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCG

N S R H A G D A V G S M A S E A E R T S W G S R G Y P H A P C S G A

ATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCC 3100
TAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGG

L S A A G V V V T R S V T A T L A S A L A P A P F A F F P S F L A

ACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGG 3200
TGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCCGTAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCC

T F A G F P R Q A L N R G I P L G F R F S A L R H L D P K K L D .

Tuesday, 18 November 1997 10:09 Page
fig 13 pCB201 (1 > 5082) Site and Sequence

GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
CACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACC 3300
G D G S R S G P S P . . T V F R P L T L E S T F F N S G L L F Q T G

AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTT
TTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACCCCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAA 3400
T T L N P I S V Y S F D L . G I L G I S A Y W L K N E L I . Q K F

AACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
TTGCGCTTAATTAAGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCA 3500
N A N . F C G M C V S . G V E S P Q A P Q A G R S M Q S M H L N .

CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCC
GTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGG 3600
S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P L T P

GCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCT
CGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGACGGAGA 3700
P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S A S

GAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGA
CTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTCGAGGGCCCTCGAACATATAGGTAAAAGCCTAGACTAGTTCTCT 3800
E L F Q K . . G G F F G G L G F C K K L P G A C I S I F G S D Q E

CAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACA
GTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGT 3900
T G . G S F R M I E Q D G L H A G S P A A W V E R L F G Y D W A Q Q

GACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTG
CTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGAC 4000
T I G C S D A A V F R L S A Q G R P V L F V K T D L S G A L N E L

CAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTAT
GTCCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATA 4100
Q D E A A R L S W L A T T G V P C A A V L D V V T E A G R D W L L

TGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGA
ACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACT 4200
L G E V P G Q D L L S S H L A P A E K V S I M A D A M R R L H T L D

TCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAA
AGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTT 4300
P A T C P F D H Q A K H R I E R A R T R M E A G L V D Q D D L D E

GAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGC
CTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACG 4400
E H Q G L A P A E L F A R L K A R M P D G E D L V V T H G D A C L

Tuesday, 18 November 1997 10:09

fig 13 pCB201 (1 > 5082) Site and Sequence Page

CGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGA
GCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACT 4500
P N I M V E N G R F S G F I D C G R L G V A D R Y Q D I A L A T R D

TATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTT
ATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAA 4600
I A E E L G G E V A D R F L V L Y G I A A P D S Q R I A F Y R L L

GACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGA
CTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACT 4700
D E F F . A G L V G S K . P T K R R P T C H H E I S I P P P P S M

AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATT
TTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAA 4800
K G V A S E S F S G T P A G . S S S A G I S C V S S S P T P T C L L

GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
CGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGT 4900
Q L I M V T N K A I A S Q I S Q I K H F F H C I L V V V C P N S S

ATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAT
TACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTA 5000
M Y L I M S V Y R R P L A R A V R N H G H S C F L C E I V I R S Q

TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
AGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAAC 5082
F H T T Y E P E A . S V K P G V P N E . A N S H . L R V

Claims

1. A vertebrate protein homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, which protein comprises an amino acid sequence having a statistically significant homology to the amino acid sequence of said UNC-53 protein of C. elegans illustrated in Figure 2.

2. A vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D or E as illustrated in Figure 9a, or block F in Figure 12a or a sequence having a statistically significant homology therewith.

3. A vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D,E or F which differ from those blocks of Figure 9a or 12a only in conservative amino acid changes.

4. A vertebrate protein having an amino acid sequence encoded by the nucleotide sequence shown from nucleotide positions 1 to 6013 illustrated in Sequence ID No. 3.

5. A vertebrate protein comprising an amino acid sequence which comprises one or more of the prosite signatures as illustrated in Figure 28 for each of said sequences of homology as claimed in claim 2.

6. A vertebrate protein comprising an amino acid

sequence as claimed in any one of claims 1 to 6 which is a human protein or a mouse protein.

7. A vertebrate protein having an amino acid sequence encoded by the nucleotide sequence shown in Sequence ID No. 4.

8. A vertebrate protein homologue according to any one of claims 1 to 7 comprising an amino acid sequence as shown in Sequence ID No. 1 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 1 in one or more conservative amino acid changes.

9. A vertebrate protein homologue according to any one of claims 1 to 7 comprising an amino acid sequence as shown in Sequence ID No. 2 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 2 in one or more conservative amino acid changes.

10. A cDNA encoding a vertebrate homologue of UNC-53 protein of <u>C. elegans</u> according to any of claims 1 to 9.

11. A cDNA according to claim 10 comprising a sequence of nucleotides encoding an amino acid sequence as shown in Sequence ID No. 1 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 1 only in one or more conservative amino acid changes.

12. A cDNA according to claim 10 comprising a sequence of nucleotides encoding an amino acid sequence as shown in Sequence ID No. 2 or an amino

acid sequence which differs from the amino acid sequence shown in Sequence ID No. 2 only in one or more conservative amino acid changes.

13. A cDNA according to any of claims 10 or 11 which cDNA comprises at least from nucleotide position 1 to position 6013 of the sequence as shown in Sequence ID No. 3.

14. A cDNA according to claim 10 or 12 which comprises the nucleotide sequence illustrated in Sequence ID No. 4.

15. A nucleic acid molecule capable of hybridising to the DNA sequences according to any of claims 10 to 14 under high stringency conditions.

16. A DNA expression vector which comprises a cDNA as claimed any of claims 10 to 14.

17. A vector according to claim 16 which comprises a promoter of C. elegans UNC-53 protein or a vertebrate homologue thereof according to any of claims 1 to 9.

18. A vector according to claim 17 wherein said promoter sequence is derived from a gene encoding a mouse or human homologue of an UNC-53 protein of C. elegans.

19. A vector according to any of claims 16 to 18 which further comprises a sequence encoding a reporter molecule.

20. A vector according to claim 19 wherein said

reporter molecule is a fluorophore.

21. A host cell transformed or transfected with the vector of any of claims 16 to 20.

22. A host cell transformed or transfected with the vector of claims 19 or 20.

23. A host cell according to claims 21 or 22, which cell comprises a prokaryotic cell such as a bacterial cell or a eukaryotic cell such as a fungal, an animal, a plant or an insect cell.

24. A transgenic cell, tissue or organism comprising a transgene capable of expressing a protein according to any of claims 1 to 9.

25. A transgenic cell, tissue or organism according to claim 24 which comprises any of a COS cell, Hep G2, MCF-7 cell, N4 mouse neuroblastoma cell, a NIH3T3 cell, or colorectal carcinoma or human derived cells.

26. A transgenic cell, tissue or organism according to claim 24 or 25 wherein said transgene comprises a vector according to any of claims 16 to 20.

27. A transgenic cell, tissue or organism according to claim 24 to 26 wherein said transgene comprises a vector according to claim 19 or 20.

28. A transgenic cell, tissue or organism according to any of claims 24 to 26 wherein said organism comprises any of an insect, a fungus, a non-

human mammal, a plant or a nematode worm.

29. A method of producing a mutant vertebrate non-human organism which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration, which method comprises inducing a mutation in the wild type gene encoding the vertebrate homologue of an UNC-53 C. elegans protein.

30. A vertebrate protein homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor thereof, for use as a medicament to promote neuronal regeneration, revascularisation, wound healing or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

31. A vertebrate protein homologue of an UNC-53 protein of C. elegans for use as claimed in claim 30 wherein said vertebrate human homologue is as claimed in any one of claims 1 to 9.

32. Use of a vertebrate protein homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor thereof, in the manufacture of a medicament for promoting neuronal regeneration, revascularisation, wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

33. Use of a vertebrate protein homologue of UNC-53 protein of C. elegans according to claim 32 wherein said vertebrate protein homologue is as

claimed in any one of claims 1 to 9.

34. A pharmaceutical composition comprising a vertebrate homologue of an UNC-53 protein of _C. elegans_, or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate protein, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

35. A pharmaceutical composition as claimed in claim 34 which comprises a vertebrate homologue of an UNC-53 protein of _C. elegans_ according to any of claims 1 to 9.

36. A nucleic acid sequence encoding a vertebrate homologue of an UNC-53 protein of _C. elegans_ or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue, for use as a medicament.

37. A nucleic acid sequence according to claim 36 wherein said sequence is a cDNA sequence as claimed in any of claims 10 to 14 or a functional fragment of said cDNA sequence.

38. Use of a nucleic acid sequence encoding a vertebrate homologue of an UNC-53 protein of _C. elegans_ or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue, in the manufacture of a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

39. Use of a nucleic acid sequence according to claim 38 wherein said sequence is a cDNA sequence as claimed in any of claims 10 to 14 or a functional fragment of said nucleic acid sequence.

40. A pharmaceutical composition comprising a nucleic acid sequence according to claim 36 or 37 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

41. A pharmaceutical composition according to claim 40 wherein said nucleic acid sequence is a cDNA sequence as claimed in any of claims 10 to 14.

42. A method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, cell shape or motility or the direction of cell migration, which method comprises contacting said compound with a host cell according to claim 21 or 23 or a transgenic cell as claimed in any of claims 24 to 27 and screening for a phenotypic change in said cell.

43. A method according to claim 41 which is capable of determining whether said compound is an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which said vertebrate homologue of an UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue is a component or is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell.

44. A method according to claim 43 wherein said

method is capable of determining whether said compound is an inhibitor or an enhancer of said vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

45. A method according to any of claims 42 to 44 wherein said phenotypic change to be screened is a change in cell growth, or shape or a change in cell motility.

46. A method according to any of claims 42 to 44 wherein said phenotypic change to be screened is a change in filopodia outgrowth, ruffling behaviour, cell adhesion, contact inhibition or the length of neurite growth.

47. A method as claimed in any of claims 42 to 44 wherein said transgenic cell is an N4 neuroblastoma cell and the phenotypic change to the screened is the length of neurite growth.

48. A method as claimed in any of claims 42 to 44 wherein said transgenic cell is an MCF-7 breast carcinoma cell or an NIH3T3 cell and the phenotypic change to be screened is the extent of phagokinesis or contact inhibition.

49. A method of determining whether a compound is an inhibitor or an enhancer of the regulation of cell shape, cell growth or motility or of the direction of cell migration, which method comprises administering said compound to a transgenic organism according to any of claims 24 to 28 or a mutant organism produced according to the method of claim 29

and screening for a phenotypic change in said organism.

50. A method according to claim 49, wherein said method is capable of determining whether said compound is an inhibitor or enhancer of a protein of the signal transduction pathway of said transgenic or mutant organisms, of which the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue is a component, or is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell.

51. A method according to claim 50 wherein said method is capable of determining whether said compound is an inhibitor or an enhancer of the vertebrate homologue of UNC-53 protein itself or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

52. A compound which is identifiable by the method according to any one of the claims 42 to 51 as an enhancer of the regulation of cell shape, or growth or motility or the direction of cell migration for use as a medicament for promoting neuronal regeneration, revascularisation or wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

53. Use of a compound which is identifiable by the method according to any one of the claims 42 to 51 as an enhancer of the regulation of cell shape, or growth or motility or the direction of cell migration in the preparation of medicament for promoting

neuronal regeneration, revascularisation or wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

54. A pharmaceutical composition comprising a compound identified according to the method of any of claims 42 to 51 claim and a pharmaceutically acceptable carrier, diluent or excipient therefor.

55. A compound which is identifiable by the method according to any one of claims 42 to 51 as an inhibitor of the regulation of cell motility, growth, or shape, or the direction of cell migration, for use as a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

56. Use of a compound according to claim 55 in the manufacture of a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

57. A pharmaceutical composition comprising the compound as claimed in claim 55, and a pharmaceutically acceptable carrier diluent or excipient therefor.

58. A method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene encoding a vertebrate homologue of UNC-53 protein of C. elegans, which method comprises the steps of (a) contacting said compound with a cell according to any of claims 22 or 27 and (b) monitoring the level of said reporter molecule and comparing the results

obtained from said monitoring step with a control comprising a cell according to claims 22 or 27, which cell has not been contacted with said compound.

59. A method as claimed in claim 58 wherein said reporter molecule detected is mRNA or green fluorescent protein.

60. A compound which is identifiable by the method according to claims 58 or 59, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, for use in promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

61. Use of a compound which is identifiable by the method of claims 58 or 59, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

62. A pharmaceutical composition which comprises the compound of claim 60 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

63. A compound which is identifiable by the method of claims 58 or 59 as an inhibitor of transcription of a gene coding for a vertebrate

homologue of a UNC-53 protein of C. elegans or a functional fragment of said gene for use in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

64. Use of a compound which is identifiable by the method of claims 58 or 59 as an inhibitor of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for alleviating spread of disease inducing cells or metastasis or loss of contact inhibition.

65. A pharmaceutical composition which comprises the compound of claim 63 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

66. A kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell motility, growth or shape or the direction of cell migration which kit comprises at least one transgenic cell as claimed in any one of claims 22 to 25 to be contacted with said compound and at least one cell according to claims 21 to 28 to be used as a control and means for contacting said compound with one of said at least one transgenic cells.

67. A kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene which kit comprises at least one cell as claimed in any one of claims 21 to 25 means for contacting said compound with said cells.

68. A kit for determining whether a compound is an enhancer or an inhibitor of the activity of a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue protein, which kit comprises at least, one vertebrate mutant non-human organism produced according to the method as claimed in claim 29 or a transgenic organism as claimed in claims 24 to 28 and a wild type of said vertebrate mutant organism.

69. A method of identifying vertebrate homologues of an unc-53 gene of C. elegans or a functional fragment thereof, which method comprises hybridizing to a DNA library a suitable oligonucleotide sequence of between 15 to 50 nucleotides of the nucleic acid sequence encoding unc-53 or a functional equivalent, derivative or bioprecursor thereof, under appropriate conditions of stringency to identify genes having statistically significant homology with the cDNA according to any of claims 10 to 14.

70. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said vertebrate homologue is a component, which method comprises:

(a) contacting an extract of said cell with an antibody to the vertebrate homologue of the UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said protein,

(b) identifying the antibody/vertebrate

homologue complex, and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein of C. elegans other than the antibody.

71. A method of identifying a further protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein or a functional equivalent, fragment or bioprecursor of said UNC-53 protein is a component, which method comprises:

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein of C. elegans in claim 70,

(b) contacting a cell extract with said antibody and identifying the antibody/protein complex,

(c) analysing the complex to identify any further protein bound to the first protein other than the antibody, and

(d) optionally repeating steps (a) to (c) to identify further proteins in said pathway.

72. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue protein is a component, which method comprises

(a) contacting an extract of said cell with the vertebrate homologue of an UNC-53 protein of C. elegans or a functional

equivalent, derivative or bioprecursor of said vertebrate homologue,
(b) identifying any vertebrate homologue of UNC-53 protein/protein complex formed and
(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

73. A method according to claim 72 which further comprises contacting a cell extract with any protein identified from step (c) not being the same as the vertebrate homologue of UNC-53 protein used and repeating steps (b) and (c) so as to identify any further protein involved in the signal transduction pathway of said cell.

74. A method of identifying a protein involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans is a component which method comprises:
(a) providing an appropriate host cell having a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain,
(b) expressing in said host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a DNA sequence according to any of claims 10 to 14 and either said DNA binding domain or the activating domain of the transcription factor,
(c) expressing in the host cell at least one second hybrid DNA sequence encoding a

putative binding protein to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion,

(d) detecting any binding of the protein being investigated with a protein according to any of claims 1 to 9 by detecting for the production of any reporter gene product in said host cell.

75. A protein identified by the method of any one of claims 70 to 74 for use as a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

76. Use of a protein identified by the methods of any one of claims 70 to 74 in the manufacture of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

77. A pharmaceutical composition comprising a protein identified by the methods of any one of claims 70 to 74 and a pharmaceutically acceptable carrier, diluent, or excipient therefor.

78. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent fragment, derivative or bioprecursor of said vertebrate homologue which process comprises culturing the cells of any of claims 21 to 28 and recovering said vertebrate homologue of

UNC-53 protein expressed.

79. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said protein which process comprises culturing an insect cell transfected with a recombinant Baculovirus vector, said vector comprising a DNA insert encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, fragment or bioprecursor of said vertebrate homologue, downstream of the Baculovirus polyhedrin promoter, and recovering the expressed vertebrate homologue of UNC-53 protein.

80. A nucleotide sequence comprising the sequence as shown in figure 15.

81. A nucleotide sequence comprising the sequence as shown in figure 16.

82. A nucleotide sequence comprising the sequence as shown in figure 17.

83. A method of detecting whether a compound is an inhibitor or an enhancer of expression of a vertebrate homologue of an UNC-53 of C. elegans, or a functional equivalent, derivative or fragment of said vertebrate homologue which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

84. A method according to claim 83 wherein said cell comprises a cell according to any of claims 21 to

28.

85. A method according to claim 83 wherein said cell has undergone loss of contact inhibition.

86. A method according to any of claims 83 to 85 which is capable of determining whether said compound is an inhibitor of expression of said vertebrate homologue in which the compound to be tested comprises a nucleic acid.

87. A method according to claim 86 wherein said nucleic acid sequence comprises an antisense DNA or RNA sequence.

88. A method according to claim 87 wherein said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

89. A method according to any of claims 83 to 85 wherein said compound to be tested comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue.

90. A method according to claim 89 wherein said protein comprises a protein identified according to any of the methods of claims 70 to 74.

91. A pharmaceutical composition comprising a compound identified according to any of claims 83 to 89 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

92. A nucleic acid sequence identified according to the method of any of claims 86 to 88 for use in

treatment of loss of contact inhibition or carcinoma which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor thereof.

93. Use of a nucleotide sequence identified according to the method of any one of claims 86 to 88 in the preparation of a medicament for the treatment of loss of contact inhibition or carcinoma which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

94. A nucleic acid according to claim 92 for use in the preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans.

95. A NIH3T3 cell line transfected with pcB201 and deposited under LMBP Accession No. 1603CB.

96. A plasmid pCB 201 of Sequence ID No. 10 deposited under LMBP Accession No. LMBP 3594.

97. A MCF-7 cell line transfected with plasmid pCB 201 deposited under LMBP Accession No. LMBP 1601CB.

98. An assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, or a functional equivalent,

derivative or bioprecursor thereof, which antibody is linked to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

99. An assay according to claim 98 wherein said reporter molecule is an antibody conjugated with a suitable fluorophore or detectable enzyme.

100. A method for detecting for expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor thereof, which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue or a functional equivalent, fragment or bioprecursor therefor with a cell extract which probe is linked to a reporter and analysing for the presence of said reporter.

101. A method according to claim 100 wherein said probe comprises a complimentary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or bioprecursor therefor.

102. A method according to claim 101 wherein said complimentary sequence is a 3' or 5' untranslated region of said mRNA.

103. A method according to claims 100 or 102 wherein said reporter comprises a radiolabel.

104. A method according to claim 100 wherein said probe comprises an antibody specific for said vertebrate homologue of said UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor therefor.

105. A method according to claim 104 wherein said reporter comprises an antibody conjugated with a detectable fluorophore or enzyme.

106. Phage Lambda clone 3b of Sequence ID No. 5 deposited under Accession No. LMBP 3595.

107. A method of determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to any of claims to 1 to 9 to microtubules or plus end regions thereof, which method comprises:

(a) contacting said compound with a transgenic cell, tissue or organism expressing UNC-53 protein or said vertebrate homologue and which protein is operably linked to a reporter molecule.

(b) screening for the localisation of said reporter molecule as compared to a cell according to step (a) which has not been contacted with said compound.

108. A compound identifiable by the method according to claim 107.

109. A compound identifiable by the method according to claim 107 as an inhibitor of localisation or association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof for

use in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

110. A compound identifiable by the method according to claim 107 as an enhancer of association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof, for use in promoting neuronal regeneration, revascularisation or wound healing, or for treating chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

111. A pharmaceutical composition comprising the compound according to claims 108 or 109 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

112. A kit for determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9 with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing UNC-53 and a reporter molecule or a cell according to any of claims 20 to 24 and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells.

113. A composition comprising UNC-53 of _C. elegans_ or a vertebrate homologue thereof according to any of claims 1 to 9 linked to a compound identified as an inhibitor or enhancer of association of UNC-53 or said vertebrate homologue with microtubules or their plus end regions for use in targeting said compound to said microtubule or the plus end regions

thereof.

114. A composition according to claim 113 which further comprises a cell transformation or transfecting agent.

115. A method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable of expressing UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting said protein to said microtubule or a plus end region thereof.

116. A method of identifying a molecule which covalently modifies UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which method comprises

a) contacting either an extract from a cell expressing UNC-53 or said vertebrate homologue or a mixture of enzymes comprising canditate UNC-53 modifying enzymes in the presence of an indicator of covalent modification of a protein,

b) identifying any covalently modified UNC-53 protein from step a),

c) identifying said molecule involved in said modification step.

117. A method according to claim 112, wherein said indicator comprises [illegible]p.

118. A method of identifying a compound which alleviates or enhances the toxicity of UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which method comprises contacting said compound with a cell, tissue or organism according to claim 27, and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof.

119. Plasmid pLM1 of Sequence ID No. 6 deposited under Accession No. LMBP 3762.

120. Plasmid pLM4 of Sequence ID No. 7 deposited under Accession No. LMBP 3763.

121. Plasmid pEGF72 of Sequence ID No. 8 deposited under Accession No. LMBP 3764.

122. Plasmid pCB501 of Sequence ID No. 9 deposited under LMBP Accession No. LMBP 3765.

123. A worm strain comprising a chimeric C.elegans human unc-53 gene deposited under LMBP Accession No. LMBP-1663CB.

124. A vertebrate homologue according to any of claims 1 to 3 which is a mouse homologue.

125. A homologue according to claim 125 having the sequence illustrated in Figure 14.

Tuesday, 18 November 1997 10:35 Page 16
fig 30 pEGFP72 (1 > 9697) Site and Sequence

```
                                                  AlwN I
                                                  |
ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9300
TGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGT
  P  R  S  A  N  P  V  T  S  G  C  C  Q  W  R  .  V  V  S  Y  R  V  G  L  K  T  I  V  T  G  .  G  A

                                         ApaL I
                                         |
GCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9400
CGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGG
  A  V  G  L  N  G  G  F  V  H  T  A  Q  L  G  A  N  D  L  H  R  T  E  I  P  T  A  .  A  M  R  K  R

ACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9500
TGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAG
  H  A  S  R  R  E  K  G  G  Q  V  S  G  K  R  Q  G  R  N  R  R  A  H  E  G  A  S  R  G  K  R  L  V  S

TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 9600
AAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCG
  L  .  S  C  R  V  S  P  P  L  T  .  A  S  I  F  V  M  L  V  R  G  A  E  P  M  E  K  R  Q  Q  R  G

                                                                                             Ava III
                                                                                             Nsi I
                                                                                             |
CTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
---------+---------+---------+---------+---------+---------+---------+---------+---------+------> 9697
GAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
  L  F  T  V  P  G  L  L  L  A  F  C  S  H  V  L  S  C  V  I  P  .  F  C  G  .  P  Y  Y  R  H  A
```

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

```
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC  100
 .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  V  P  A  V  L  T

CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA  200
 A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S  N  A  N  R  D  F  P  L  T  S  M  G  G  V  F  T  V

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT  300
 N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A  P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V

CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT  400
 H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H  R  Y  Y  H  G  D  A  V  L  A  V  H  Q  V  A  V

TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT  500
 I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q  V  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .

ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT  600
 Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G  G  L  Y  K  Q  S  V  F  S  E  P  S  D  P  L  A  L

CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC  700
                                         eGFPC.e.unc53sma
 P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG  800
                                         eGFPC.e.unc53sma
 V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC  900
                                         eGFPC.e.unc53sma
 T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC  1000
                                         eGFPC.e.unc53sma
 R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I
```

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 2

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT 1100

eGFPC.e.unc53sma

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I I

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 1200

eGFPC e.unc53sma

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT 1300

eGFPC.e.unc53sma

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG 1400

eGFPC.e.unc53sma

C.e.unc53 sma

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT 1500

eGFPC.e.unc53sma

C.e.unc53 sma

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTGTAGCTTGGACCTACCGGAGCTTTGCACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA 1600

eGFPC.e.unc53sma

C.e.unc53 sma

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

CGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCGTCTTCGAAGCCGTTGACTT 1700

eGFPC.e.unc53sma

C.e.unc53 sma

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L I

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence Page 1

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCA
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT 1800

eGFPC.e.unc53sma

C.e.unc53 sma

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT 1900

eGFPC.e.unc53sma

C.e.unc53 sma

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT 2000

eGFPC.e.unc53sma

C.e.unc53 sma

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTTGGCTGGATGGAGGGTTGAGGTTTTT 2100

eGFPC.e.unc53sma

C.e.unc53 sma

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q K

CCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTG
GGAAGATCTGGTGTTTGGGTCGATCAAGCACAACGATGTTGATGTTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGGTTTTGAAC 2200

eGFPC.e.unc53sma

C.e.unc53 sma

P S R P Q T Q L V R V A T T T K I G S S K L A A P K A V S T P K L

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCATC
GAAGACACTTCTGATAACCTCGTTTTGTTCTCGGGCTATTGTCGCCACCACCACCACCACCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG 2300

eGFPC.e.unc53sma

C.e.unc53 sma

A S V K T I G A K Q E P D N S G G G G G G M L K L K L F S S K N P S

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence Page 4

TTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG
AAGGAGTAGCTTATCGGGTGTTGGATGCTCTTTCCGCCGCCGCCACGGAGTTGTTGTTTGAAACAGCTTTTAGCGACGGGGTCACTTTTCACCGGACTTC 2400

eGFPC.e.unc53sma

C.e.unc53 sma

S S S N S P Q P T R K A A A V P Q Q Q T L S K I A A P V K S G L H

CCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAG
GGCGGCTGGTCATTCGACCCTTCACGGTGCAGATACAGCTTCGAAACATGCGGTTTTCAAAGGATGGCATTTTGCCTGCGGGGTTAGTATAGAGTTGTTC 2500

eGFPC.e.unc53sma

C.e.unc53 sma

P P T S K L G S A T S M S K L C T P K V S Y R K T D A P I I S Q Q

ACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTTCTCAGGCCTATGCGACCTAAGTTGTCGTGCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTCGTA 2600

eGFPC.e.unc53sma

C.e.unc53 sma

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTCGATAGTCTGTCGGC 2700

eGFPC.e.unc53sma

C.e.unc53 sma

H S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTCGGACAACTCCTTTTTGGTTGTGACCGTCACTTTCCTCACTTTTCGTGTCGCTTTTTTCTAG 2800

eGFPC.e.unc53sma

C.e.unc53 sma

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTCGACAAGGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGTATTCTTCAACTGTTTACTGGGGCACTATAGACT 2900

eGFPC.e.unc53sma

C.e.unc53 sma

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence


```
AAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA
TTTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAGT  3000
                                    eGFPC.e.unc53sma
                                    C.e.unc53 sma
 K  P  E  P  E  K  L  Q  S  M  S  I  D  T  T  D  V  P  P  L  P  P  L  K  S  V  V  P  L  K  M  T  S

ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACT
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTTAGTGTAGCGGACAGTTCAGCAAACCTATACTCGTCAGCAGGCGCAGACTTCTGA  3100
                                    eGFPC.e.unc53sma
                                    C.e.unc53 sma
 I  R  Q  P  P  T  Y  D  V  L  L  K  Q  G  K  I  T  S  P  V  K  S  F  G  Y  E  Q  S  S  A  S  E  D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTTTGAAGACCATTAGTAAGCGACCTCTCTTCCTACCCTTTCTTATTCTGTAGTCTTAG  3200
                                    eGFPC.e.unc53sma
                                    C.e.unc53 sma
 S  I  V  A  H  A  S  A  Q  V  T  P  P  T  K  T  S  G  N  H  S  L  E  R  R  M  G  K  N  K  T  S  E  S

CAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTTACTCCCTCTTCGACTTTCTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA  3300
                                    eGFPC.e.unc53sma
                                    C.e.unc53 sma
 S  G  Y  T  S  D  A  G  V  A  M  C  A  K  M  R  E  K  L  K  E  Y  D  D  M  T  R  R  A  Q  N  G  Y

CCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGG
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCCTGCTAAACAGGCCTCATCTGTACC  3400
                                    eGFPC.e.unc53sma
                                    C.e.unc53 sma
 P  D  N  F  E  D  S  S  S  L  S  S  G  I  S  D  N  N  E  L  D  D  I  S  T  D  D  L  S  G  V  D  M

CAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTCGGTCCTCCACATCAGT
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAAACAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA  3500
                                    eGFPC.e.unc53sma
                                    C.e.unc53 sma
 A  T  V  A  S  K  H  S  D  Y  S  H  F  V  R  H  P  T  S  S  S  S  K  P  R  V  P  S  R  S  S  T  S  V
```

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 6

CGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGGACAA
GCTAAGAGCTAGAGCTCGTCTTGTCCTCTTACACATGTTTGAAGACAGGGTCACGGCTTGCTCGGTTGCACCGCGGCGACGGTGGAGTTGGAAGCCTGTT 3600

eGFPC.e.unc53sma

C.e.unc53 sma

D S R S R A E Q E N V Y K L L S Q C R T S Q R G A A A T S T F G Q

CATTCGCTAAGATCCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCAC
GTAAGCGATTCTAGGGGCCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTG 3700

eGFPC.e.unc53sma

C.e.unc53 sma

H S L R S P G S T G S R . L I I I S H T T F V E V L L A L K N L P

ACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
TGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTT 3800
H L P L N L K H K M N A I V V V N L F I A A Y N G Y K . S N S I T N

TTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTT
AAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAACATTCGCAATTATAAAACAA 3900
F T N K A F F S L H S S C G L S K L I N V S . R V N C K R . Y F V

AAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGG
TTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCC 4000
K I R V K F L L N Q L I F . P I G R N R Q N P L . I K R I D R D R

TTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTAC
AACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATG 4100
V E C C S S L E Q E S T I K E R G L Q R Q R A K N R L S G R W P T T

GTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAA
CACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTT 4200
. T I T L I K F F G V E V P . S T K S E P . R E P P I . S L T G K

GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCC
CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGG 4300
A G E R G E K G R E E S E R S G R . G A G K C S G H A A R N H H T

GCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATA
CGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTAT 4400
R R A . C A A T G R V R W H F S G K C A R N P Y L F I F L N T F K Y

TGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAG
ACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAGGACTCCGCCTTTCTTGGTCGACACCTTACACACAGTC 4500
V S A H E T I T L I N A S I I L K K E E S . G G K N Q L W N V C Q

```
TTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCC
AATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGG 4600
 L  G  C  G  K  S  P  G  S  P  A  G  R  S  M  Q  S  M  H  L  N  .  S  A  T  R  C  G  K  S  P  G  S

CAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGC
GTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCG 4700
 P  A  G  R  S  M  Q  S  M  H  L  N  .  S  A  T  I  V  P  P  L  T  P  P  I  P  P  L  T  P  P  S  S  A

CCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTT
GGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAA 4800
 H  S  P  P  H  G  .  L  I  F  F  I  Y  A  E  A  E  A  A  S  A  S  E  L  F  Q  K  .  .  G  G  F  F  .

GGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGG
CCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACC 4900
 G  G  L  G  F  C  K  D  R  S  R  D  R  M  R  I  V  S  H  D  .  T  R  W  I  A  R  R  F  S  G  R  L

GTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTG
CACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAAC 5000
 G  G  E  A  I  R  L  .  L  G  T  T  D  N  R  L  L  .  C  R  R  V  P  A  V  S  A  G  A  P  G  S  F  C

TCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGA
AGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCT 5100
 Q  D  R  P  V  R  C  P  E  .  T  A  R  R  G  S  A  A  I  V  A  G  H  D  G  R  S  L  R  S  C  A  R

CGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATC
GCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAG 5200
 R  C  H  .  S  G  K  G  L  A  A  I  G  R  S  A  G  A  G  S  P  V  I  S  P  C  S  C  R  E  S  I  H

ATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGG
TACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACC 5300
 H  G  .  C  N  A  A  A  A  Y  A  .  S  G  Y  L  P  I  R  P  P  S  E  T  S  H  R  A  S  T  Y  S  D  G

AAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGA
TTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCCGCT 5400
 S  R  S  C  R  S  G  .  S  G  R  R  A  S  G  A  R  A  S  R  T  V  R  Q  A  Q  G  E  H  A  R  R  R

GGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCG
CCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGC 5500
 G  S  R  R  D  P  W  R  C  L  L  A  E  Y  H  G  G  K  W  P  L  F  W  I  H  R  L  W  P  A  G  C  G

GACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAG 5600
 G  P  L  S  G  H  S  V  G  Y  P  .  Y  C  .  R  A  W  R  R  M  G  .  P  L  P  R  A  L  R  Y  R  R  S

CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCA
GGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGT 5700
 R  F  A  A  H  R  L  L  S  P  S  .  R  V  L  L  S  G  T  L  G  F  E  M  T  D  Q  A  T  P  N  L  P
```

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence


```
TCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCA
AGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGT 5800
 S  R  D  F  D  S  T  A  A  F  Y  E  R  L  G  F  G  I  V  F  R  D  A  G  W  M  I  L  Q  R  G  D  L

TGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAA
ACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATGGCCTTCCTTGGGCGCGATACTGCCGTTATTTTTCTGTCTT 5900
M  L  E  F  F  A  H  P  R  G  R  L  T  E  T  R  K  E  T  I  P  E  G  T  R  A  M  T  A  I  K  R  Q  N

TAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATAC
ATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATG 6000
 K  T  H  G  V  G  S  F  V  H  K  R  G  V  R  S  Q  G  W  H  S  V  D  T  P  P  R  P  H  W  G  Q  Y  .

GCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAG
CGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTCGGTTGCAGCCCCGCCGTCCGGGACGGTATCGGAGTC 6100
 A  R  V  S  S  F  S  P  P  H  P  P  S  S  G  E  G  P  G  L  A  A  N  V  G  A  A  G  P  A  I  A  S

GTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTA
CAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAAT 6200
G  Y  S  Y  I  L  .  I  D  L  K  L  H  F  .  F  K  R  I  .  V  K  I  L  F  D  N  L  M  T  K  I  P  .

ACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACA
TGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGT 6300
 R  E  F  S  F  H  .  A  S  D  P  V  E  K  I  K  G  S  S  .  D  P  F  F  L  R  V  I  C  C  L  Q  T

AAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATA
TTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTAT 6400
K  K  P  P  L  P  A  V  V  C  L  P  D  Q  E  L  P  T  L  F  P  K  V  T  G  F  S  R  A  Q  I  P  N

CTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC
GACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACG 6500
T  V  L  L  V  .  P  .  L  G  H  H  F  K  N  S  V  A  P  P  T  Y  L  A  L  L  I  L  L  P  V  A  A  A

CAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCC
GTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGG 6600
 S  G  D  K  S  C  L  T  G  L  D  S  R  R  .  L  P  D  K  A  Q  R  S  G  .  T  G  G  S  C  T  Q  P

AGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCC 6700
S  L  E  R  T  T  Y  T  E  L  R  Y  L  Q  R  E  L  .  E  S  A  T  L  P  E  G  R  K  A  D  R  Y  P

TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGA
ATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACT 6800
V  S  G  R  V  G  T  G  E  R  T  R  E  L  P  G  G  N  A  V  Y  L  Y  S  P  V  G  F  R  H  L  .  L  E

GCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGA 6900
 R  R  F  L  .  C  S  S  G  G  R  S  L  V  K  N  A  S  N  A  A  F  L  R  F  L  A  F  C  W  P  F  A
```

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 9

```
CACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
------------------------------------------------------------> 6960
GTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
 H  M  F  F  P  A  L  S  P  D  S  V  D  N  R  I  T  A  M  H
------------------------------------------------------------>
```

6

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence 3p. Page I
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

```
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC 100
 . L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA 200
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT 300
 N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT 400
 H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W

TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT 500
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .

ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT 600
 Q L R P I D A N G R . A C T V G G L Y K Q S W F S E P S D P L A L

CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC 700
                                   eGFPC.e.unc53ecl
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG 800
                                   eGFPC.e.unc53ecl
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC 900
                                   eGFPC.e.unc53ecl
 T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 1000
                                   eGFPC.e.unc53ecl
R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I
```

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence


```
ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT  1100
                              eGFPC.e.unc53ecl
D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC  1200
                              eGFPC.e.unc53ecl
 V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT  1300
                              eGFPC.e.unc53ecl
P  D  N  H  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG  1400
                              eGFPC.e.unc53ecl
                                          C.e.unc53 ecl
I  T  L  G  M  D  E  L  Y  K  S  G  L  R  S  T  S  N  V  E  L  I  P  I  Y  T  D  W  A  N  R  H  L  S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT  1500
                              eGFPC.e.unc53ecl
                              C.e.unc53 ecl
 K  G  S  L  S  K  S  I  R  D  I  S  N  D  F  R  D  Y  R  L  V  S  Q  L  I  N  V  I  V  P  I  N  E

TTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTGTAGCTTGGACCTACCGGAGCTTTGCACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA  1600
                              eGFPC.e.unc53ecl
                              C.e.unc53 ecl
F  S  P  A  F  T  K  R  L  A  K  I  T  S  N  L  D  G  L  E  T  C  L  D  Y  L  K  N  L  G  L  D  C

CGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCGTCTTCGAAGCCGTTGACTT  1700
                              eGFPC.e.unc53ecl
                              C.e.unc53 ecl
S  K  L  T  K  T  D  I  D  S  G  N  L  G  A  V  L  Q  L  L  F  L  L  S  T  Y  K  Q  K  L  R  Q  L  I
```

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence Page 3

```
AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCA
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT 1800
eGFPC.e.unc53ecl
C.e.unc53 ecl
 K  D  Q  K  K  L  E  Q  L  P  T  S  I  M  P  P  A  V  S  K  L  P  S  P  R  V  A  T  S  A  T  A  S

GCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT 1900
eGFPC.e.unc53ecl
C.e.unc53 ecl
 A  T  N  P  N  S  N  F  P  Q  M  S  T  S  R  L  Q  T  P  Q  S  R  I  S  K  I  D  S  S  K  I  G  I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT 2000
eGFPC.e.unc53ecl
C.e.unc53 ecl
K  P  K  T  S  G  L  K  P  P  S  S  S  T  T  S  S  N  N  T  N  S  F  R  P  S  S  R  S  S  G  N  N  N

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTTGGCTGGATGGAGGGTTGAGGTTTTT 2100
eGFPC.e.unc53ecl
C.e.unc53 ecl
 V  G  S  T  I  S  T  S  A  K  S  L  E  S  S  S  T  Y  S  S  I  S  N  L  N  R  P  T  S  Q  L  Q  K

CCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTG
GGAAGATCTGGTGTTTGGGTCGATCAAGCACAACGATGTTGATGTTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGGTTTTGAAC 2200
eGFPC.e.unc53ecl
C.e.unc53 ecl
 P  S  R  P  Q  T  Q  L  V  R  V  A  T  T  T  K  I  G  S  S  K  L  A  A  P  K  A  V  S  T  P  K  L

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCATC
GAAGACACTTCTGATAACCTCGTTTTGTTCTCGGGCTATTGTCGCCACCACCACCACCACCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG 2300
eGFPC.e.unc53ecl
C.e.unc53 ecl
A  S  V  K  T  I  G  A  K  Q  E  P  D  N  S  G  G  G  G  G  G  M  L  K  L  K  L  F  S  S  K  N  P  S
```

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence

Page 4

TTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG
AAGGAGTAGCTTATCGGGTGTTGGATGCTCTTTCCGCCGCCGCCACGGAGTTGTTGTTTGAAACAGCTTTTAGCGACGGGGTCACTTTTCACCGGACTTC 2400

eGFPC.e.unc53ecl

C.e.unc53 ecl

S S S N S P Q P T R K A A A V P Q Q Q T L S K I A A P V K S G L K

CCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAG
GGCGGCTGGTCATTCGACCCTTCACGGTGCAGATACAGCTTCGAAACATGCGGTTTTCAAAGGATGGCATTTTGCCTGCGGGGTTAGTATAGAGTTGTTC 2500

eGFPC.e.unc53ecl

C.e.unc53 ecl

P P T S K L G S A T S M S K L C T P K V S Y R K T D A P I I S Q Q

ACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTTCTCAGGCCTATGCGACCTAAGTTGTCGTGCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTCGTA 2600

eGFPC.e.unc53ecl

C.e.unc53 ecl

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTCGATAGTCTGTCGGC 2700

eGFPC.e.unc53ecl

C.e.unc53 ecl

H S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTCGGACAACTCCTTTTTGGTTGTGACCGTCACTTTCCTCACTTTTCGTGTCGCTTTTTTCTAG 2800

eGFPC.e.unc53ecl

C.e.unc53 ecl

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTCGACAAGGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGTATTCTTCAACTGTTTACTGGGGCACTATAGACT 2900

eGFPC.e.unc53ecl

C.e.unc53 ecl

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence
Page 6

```
AAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA
TTTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAGT 3000
                                 eGFPC.e.unc53ecl
                                 C.e.unc53 ecl
 K  P  E  P  E  K  L  Q  S  M  S  I  D  T  T  D  V  P  P  L  P  P  L  K  S  V  V  P  L  K  M  T  S

ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACT
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTTAGTGTAGCGGACAGTTCAGCAAACCTATACTCGTCAGCAGGCGCAGACTTCTGA 3100
                                 eGFPC.e.unc53ecl
                                 C.e.unc53 ecl
 I  R  Q  P  P  T  Y  D  V  L  L  K  Q  G  K  I  T  S  P  V  K  S  F  G  Y  E  Q  S  S  A  S  E  D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTTTGAAGACCATTAGTAAGCGACCTCTCTTCCTACCCTTTCTTATTCTGTAGTCTTAG 3200
                                 eGFPC.e.unc53ecl
                                 C.e.unc53 ecl
 S  I  V  A  H  A  S  A  Q  V  T  P  P  T  K  T  S  G  N  H  S  L  E  R  R  M  G  K  N  K  T  S  E  S

CAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTTACTCCCTCTTCGACTTTCTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA 3300
                                 eGFPC.e.unc53ecl
                                 C.e.unc53 ecl
 S  G  Y  T  S  D  A  G  V  A  M  C  A  K  M  R  E  K  L  K  E  Y  D  D  M  T  R  R  A  Q  N  G  Y

CCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGAGGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACC
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCCCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGG 3400
                       eGFPC.e.unc53ecl
                  C.e.unc53 ecl
 P  D  N  F  E  D  S  S  S  L  S  S  G  I  S  D  N  N  E  G  I  H  R  I  .  I  T  D  H  N  Q  P  Y

ACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTG
TGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAAC 3500
 H  I  C  R  G  F  T  C  F  K  K  P  P  T  P  P  P  E  P  E  T  .  N  E  C  N  C  C  C  .  L  V  Y  C

CAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAA
GTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTT 3600
 S  L  .  W  L  Q  I  K  Q  .  H  H  K  F  H  K  .  S  I  F  F  T  A  F  .  L  W  F  V  Q  T  H  Q
```

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence


```
TGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGG
ACATAGAATTGCGCATTTAACATTCGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCC 3700
 C  I  L  T  R  K  L  .  A  L  I  F  C  .  N  S  R  .  I  F  V  K  S  A  H  F  L  T  N  R  P  K  S

CAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTC
GTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAG 3800
 A  K  S  L  I  N  Q  K  N  R  P  R  .  G  .  V  L  F  Q  F  G  T  R  V  H  Y  .  R  T  W  T  P  T  S

AAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGA
TTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCT 3900
 K  G  E  K  P  S  I  R  A  M  A  H  Y  V  N  H  H  P  N  Q  V  F  W  G  R  G  A  V  K  H  .  I  G

ACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCT
TGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGA 4000
 T  L  K  G  A  P  D  L  E  L  D  G  E  S  R  R  T  W  R  E  R  K  G  R  K  R  K  E  R  A  L  G  R

GGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGC
CCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCG 4100
 W  Q  V  .  R  S  R  C  A  .  P  P  H  P  P  R  L  M  R  R  Y  R  A  R  Q  V  A  L  F  G  E  M  C  A

GGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
CCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCT 4200
 E  P  L  F  V  Y  F  S  K  Y  I  Q  I  C  I  R  S  .  D  N  N  P  D  K  C  F  N  N  I  E  K  G  R

GTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTC
CAGGACTCCGCCTTTCTTGGTCGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAG 4300
 V  L  R  R  K  E  P  A  V  E  C  V  S  V  R  V  W  K  V  P  R  L  P  S  R  Q  K  Y  A  K  H  A  S

AATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
TTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGG 4400
 Q  L  V  S  N  Q  V  W  K  V  P  R  L  P  S  R  Q  K  Y  A  K  H  A  S  Q  L  V  S  N  H  S  P  A  P

TAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTC
ATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAG 4500
 N  S  A  H  P  A  P  N  S  A  Q  F  R  P  F  S  A  P  W  L  T  N  F  F  Y  L  C  R  G  R  G  R  L

GGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT
CCGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTA 4600
 G  L  .  A  I  P  E  V  V  R  R  L  F  W  R  P  R  L  L  Q  R  S  I  K  R  Q  D  E  D  R  F  A  .

TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCC
ACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGG 4700
 L  N  K  M  D  C  T  Q  V  L  R  P  L  G  W  R  G  Y  S  A  M  T  G  H  N  R  Q  S  A  A  L  M  P  P

GTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGT
CACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCA 4800
 C  S  G  C  Q  R  R  G  A  R  F  F  L  S  R  P  T  C  P  V  P  .  M  N  C  K  T  R  Q  R  G  Y  R
```

```
GGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
CCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGA 4900
 G  W  P  R  R  A  F  L  A  Q  L  C  S  T  L  S  L  K  R  E  G  T  G  C  Y  V  A  K  C  R  G  R  I

CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCAC
GGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTG 5000
 S  C  H  L  T  L  L  L  P  R  K  Y  P  S  V  L  M  Q  C  G  G  C  I  R  L  I  R  L  P  A  H  S  T  T

CAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCG
GTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGC 5100
 K  R  N  I  A  S  S  E  H  V  L  G  W  K  P  V  L  S  I  R  M  I  V  T  K  S  I  R  G  S  R  Q  P

AACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCG
TTGACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGC 5200
 N  C  S  P  G  S  R  R  A  C  P  T  A  R  I  S  S  .  P  M  A  M  P  A  C  R  I  S  W  W  K  M  A

CTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAA
GAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTT 5300
 A  F  L  D  S  S  T  V  A  G  W  V  W  R  T  A  I  R  T  .  R  W  L  P  V  I  L  L  K  S  L  A  A  N

TGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
ACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGA 5400
 G  L  T  A  S  S  C  F  T  V  S  P  L  P  I  R  S  A  S  P  S  I  A  F  L  T  S  S  S  E  R  D  S

GGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTC
CCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAG 5500
 G  V  R  N  D  R  P  S  D  A  Q  P  A  I  T  R  F  R  F  H  R  R  L  L  .  K  V  G  L  R  N  R  F

CGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
GCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATG 5600
 P  G  R  R  L  D  D  P  P  A  R  G  S  H  A  G  V  L  R  P  P  .  G  E  A  N  .  N  T  E  G  D  N  T

CGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCA
GCCTTCCTTGGGCGCGATACTGCCGTTATTTTTCTGTCTTATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGT 5700
 G  R  N  P  R  Y  D  G  N  K  K  T  E  .  N  A  R  C  V  V  V  C  S  .  T  R  G  S  V  P  G  L  A

CTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTC
GAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAG 5800
 L  C  R  Y  P  T  E  T  P  L  G  P  I  R  P  R  F  F  L  F  P  T  P  P  P  K  F  G  .  R  P  R  A

GCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGG
CGTCGGTTGCAGCCCCGCCGTCCGGGACGGTATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCC 5900
 R  S  Q  R  R  G  G  R  P  C  H  S  L  R  L  L  I  Y  T  L  D  .  F  K  T  S  F  L  I  .  K  D  L  G

TGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
ACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAAC 6000
 E  D  P  F  .  .  S  H  D  Q  N  P  L  T  .  V  F  V  P  L  S  V  R  P  R  R  K  D  Q  R  I  F  L
```

Tuesday, 18 November 1997 11:46
fig 32 pEGFPecl (1 > 6700) Site and Sequence


```
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTT
TCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAA 6100
 R  S  F  F  S  A  R  N  L  L  L  A  N  K  K  T  T  A  T  S  G  G  L  F  A  G  S  R  A  T  N  S  F

CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTA
GGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGAT 6200
S  E  G  N  W  L  Q  Q  S  A  D  T  K  Y  C  P  S  S  V  A  V  V  R  P  P  L  Q  E  L  C  S  T  A  Y

CATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCG 6300
  I  P  R  S  A  N  P  V  T  S  G  C  C  Q  W  R  .  V  V  S  Y  R  V  G  L  K  T  I  V  T  G  .  G

GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGC
CGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCG 6400
 A  A  V  G  L  N  G  G  F  V  H  T  A  Q  L  G  A  N  D  L  H  R  T  E  I  P  T  A  .  A  M  R  K

GCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGT
CGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCA 6500
R  H  A  S  R  R  E  K  G  G  Q  V  S  G  K  R  Q  G  R  N  R  R  A  H  E  G  A  S  R  G  K  R  L  V

ATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
TAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCG 6600
  S  L  .  S  C  R  V  S  P  P  L  T  .  A  S  I  F  V  M  L  V  R  G  A  E  P  M  E  K  R  Q  Q  R

GGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 6700
 G  L  F  T  V  P  G  L  L  L  A  F  C  S  H  V  L  S  C  V  I  P  .  F  C  G  .  P  Y  Y  R  H  A
```

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

```
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC 100
 .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  V  P  A  V  L  T

CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA 200
A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S  N  A  N  R  D  F  P  L  T  S  M  G  G  V  F  T  V

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT 300
 N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A  P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V

CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT 400
 H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H  R  Y  Y  H  G  D  A  V  L  A  V  H  Q  V  A  V

TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT 500
I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q  V  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .

ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT 600
 Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G  G  L  Y  K  Q  S  V  F  S  E  P  S  D  P  L  A  L

CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC 700
                                    eGFPC.e.unc53xba
P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG 800
                                    eGFPC.e.unc53xba
V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC 900
                                    eGFPC.e.unc53xba
 T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 1000
                                    eGFPC.e.unc53xba
R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I
```

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence Page 4

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT 1100
eGFPC.e.unc53xba
D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 1200
eGFPC.e.unc53xba
V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

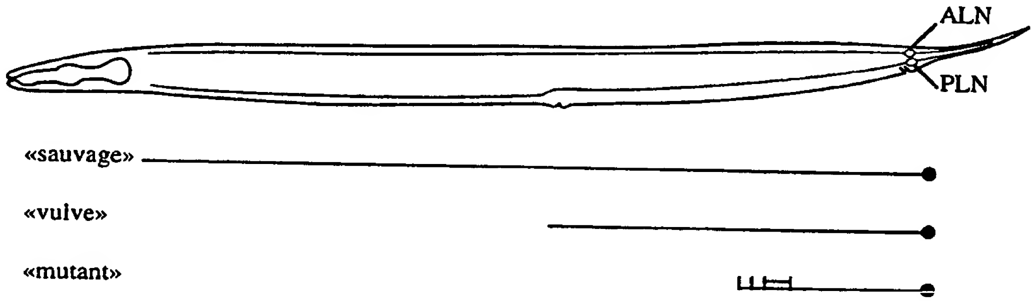
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT 1300
eGFPC.e.unc53xba
P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG 1400
eGFPC.e.unc53xba
C.e.unc53 xba
I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT 1500
eGFPC.e.unc53xba
C.e.unc53 xba
K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTGTAGCTTGGACCTACCGGAGCTTTGCACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA 1600
eGFPC.e.unc53xba
C.e.unc53 xba
F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

CGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCGTCTTCGAAGCCGTTGACTT 1700
eGFPC.e.unc53xba
C.e.unc53 xba
S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L K



souches	phénotypes	sauvage	vulve	mutant	nombre
wt	ALN	100	0	0	70
	PLN	100	0	0	70
unc-53(n152)	ALN	0	26	74	50
	PLN	nd	nd	nd	
unc-53(n152) pA*unc-53*	ALN	95	0	5	65
	PLN	77	0	23	44
unc-53(n152) pA*unc-53-H1*	ALN	25	60	15	56
	PLN	52	16	32	25

Figure 51b

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence
Enzymes : All 146 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA 100
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACTGGCTGTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGC 200
G G R . V V R E Q N L S Y N K A R L D R Q L H E E S A . G . A F C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGGTCTATATGCGCAACTGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT 300
A A S R C T G Q I Y A L T L I I D . L L I V I N Y G V I S S . P I Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCA 400
G V P R Y I T Y G K W P A W L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG 500
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT 600
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTT 700
R Y Y H G D A V L A V H Q W A W I A V . L T G I S K S P P H . R Q

TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG
ACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTC 800
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG 900
T7 promotor priming site
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S W L A

GTTTAAACTTAAGCTTACCATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG 1000
ProBond binding domain
F K L K L T M G G S H H H H H H G M A S M T G G Q Q M G R D L Y D

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence


```
GATGACGATAAGGTACCTAGGATCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGA
CTACTGCTATTCCATGGATCCTAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCCTTTACT  100
                                         pLM5 insert = U3
 D  D  D  K  V  P  R  I  H  A  N  E  E  E  E  P  E  K  K  E  V  S  E  L  R  S  E  L  W  E  K  E  M

AGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAA
TCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTT  200
                                         pLM5 insert = U3
                                              ORF U3
 K  L  T  D  I  R  L  E  A  L  N  S  A  H  Q  L  D  Q  L  R  E  T  M  H  N  M  Q  L  E  V  D  L  L  K

AGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCC
TCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGG  300
                                         pLM5 insert = U3
                                              ORF U3
 A  E  N  D  R  L  K  V  A  P  G  P  S  S  G  S  T  P  G  Q  V  P  G  S  S  A  L  S  S  P  R  R  S

CTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGA
GATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACT  400
                                         pLM5 insert = U3
                                              ORF U3
 L  G  L  A  L  T  H  S  F  G  P  S  L  A  D  T  D  L  S  P  M  D  G  I  S  T  C  G  P  K  E  E  V

CCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGT
GGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCA  500
                                         pLM5 insert = U3
                                              ORF U3
 T  L  R  V  V  V  R  M  P  P  Q  H  I  I  K  G  D  L  K  Q  Q  E  F  F  L  G  C  S  K  V  S  G  K  V

TGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATC
ACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAG  600
                                         pLM5 insert = U3
                                              ORF U3
 D  V  K  M  L  D  E  A  V  F  Q  V  F  K  D  Y  I  S  K  M  D  P  A  S  T  L  G  L  S  T  E  S  I
```

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 3

```
CATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCA
GTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAG 700
pLM5 insert = U3
ORF U3
 H  G  Y  S  I  S  H  V  K  R  V  L  D  A  E  P  P  E  M  P  P  C  R  R  G  V  N  N  I  S  V  S  L

AAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCG
TTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGC 1800
pLM5 insert = U3
ORF U3
 K  G  L  K  E  K  C  V  D  S  L  V  F  E  T  L  I  P  K  P  M  M  Q  H  Y  I  S  L  L  L  K  H  R  R

CCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGC
GGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCG 1900
pLM5 insert = U3
ORF U3
 L  V  L  S  G  P  S  G  T  G  K  T  Y  L  T  N  R  L  A  E  Y  L  V  E  R  S  G  R  E  V  T  E  G

ATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGG
TAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCC 2000
pLM5 insert = U3
ORF U3
 I  V  S  T  F  N  M  H  Q  Q  S  C  K  D  L  Q  L  Y  L  S  N  L  A  N  Q  I  D  R  E  T  G  I  G

ATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTA
TACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGAT 2100
pLM5 insert = U3
ORF U3
 D  V  P  L  V  I  L  L  D  D  L  S  E  A  G  S  I  S  E  L  V  N  G  A  L  T  C  K  Y  H  K  C  P  Y

TATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCC
ATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGG 2200
pLM5 insert = U3
ORF U3
 I  I  G  T  T  N  Q  P  V  K  M  T  P  N  H  G  L  H  L  S  F  R  M  L  T  F  S  N  N  V  E  P  A
```

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 4

AATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTAC
TTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATG 2300

pLM5 insert = U3

ORF U3

N G F L V R Y L R R K L V E S D S D I N A N K E E L L R V L D W V

CCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGA
GGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACT 2400

pLM5 insert = U3

ORF U3

P K L W Y H L H T F L E K H S T S D F L I G P C F F L S C P I G I E

GGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCT
CCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGA 2500

pLM5 insert = U3

ORF U3

D F R T W F I D L W N N S I I P Y L Q E G A K D G I K V H G Q K A

GCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGG
CGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACC 2600

pLM5 insert = U3

ORF U3

A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H L P P P T V

GCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACT
CGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGA 2700

pLM5 insert = U3

ORF U3

G P H S I A S P P E D R T V K D S T P S S L D S D P L M A M L L K L

TCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCC
AGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGG 2800

pLM5 insert = U3

ORF U3

Q E A A N Y I E S P D R E T I L D P N L Q A T L . G F G N H C H P

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 5

```
CGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAG
GCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTC 2900
pLM5 insert = U3
R  T  A  E  R  V  H  Q  L  S  .  L  L  L  S  P  L  L  F  Q  S  T  G  S  P  A  P  G  G  E  Q  E  G

GAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAA
CTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTT 3000
pLM5 insert = U3
G  G  D  E  R  G  G  T  G  S  V  C  C  T  F  E  N  F  L  G  R  N  G  G  V  A  F  G  N  L  C  P  L  N

CACATTTACTGGCCTCCTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GTGTAAATGACCGGAGGAGATCTCCCGGGCAAATTTGGGCGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCA 3100
pLM5 insert = U3
T  F  T  G  L  L  .  R  A  R  L  N  P  L  I  S  L  D  C  A  F  .  L  P  A  I  C  C  L  P  L  P  R

GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
CGGAAGGAACTGGGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTTACTCCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCC 3200
A  F  L  D  P  G  R  C  H  S  H  C  P  F  L  I  K  .  G  N  C  I  A  L  S  E  .  V  S  F  Y  S  G

GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCA
CCACCCCACCCCGTCCTGTCGTTCCCCCTCCTAACCCTTCTGTTATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCCTTTCTTGGT 3300
G  V  G  G  A  G  Q  Q  G  G  G  L  G  R  Q  .  Q  A  C  V  G  C  G  G  L  Y  G  F  .  G  G  K  N  Q

GCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGC
CGACCCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCGTAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCG 3400
L  G  L  .  G  V  S  P  R  A  L  .  R  R  I  K  R  G  G  C  G  G  Y  A  Q  R  D  R  Y  T  C  Q  R

CCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGA
GGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCGTAGGGAAATCCCAAGGCT 3500
P  S  A  R  S  F  R  F  L  P  F  L  S  R  H  V  R  R  L  S  P  S  S  S  K  S  G  H  P  F  R  V  P

TTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT
AAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCA 3600
I  .  C  F  T  A  P  R  P  Q  K  T  .  L  G  .  W  F  T  .  W  A  I  A  L  I  D  G  F  S  P  F  D  V

TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGAT
ACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACCTTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACCCCTA 3700
G  V  H  V  L  .  .  V  T  L  V  P  N  W  N  N  T  Q  P  Y  L  G  L  F  F  .  F  I  R  D  F  G  D

TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCT
AAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATTGCGCTTAATTAAGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGA 3800
F  G  L  L  V  K  K  .  A  D  L  T  K  I  .  R  E  L  I  L  W  N  V  C  Q  L  G  C  G  K  S  P  G
```

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence Page 6

```
CCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
GGGGTCCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGT 3900
 S  P  G  R  Q  K  Y  A  K  H  A  S  Q  L  V  S  N  Q  V  V  K  V  P  R  L  P  S  R  Q  K  Y  A  K  H

TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT
ACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGA 4000
  A  S  Q  L  V  S  N  H  S  P  A  P  N  S  A  H  P  A  P  N  S  A  Q  F  R  P  F  S  A  P  V  L  T

AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAG
TTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTC 4100
 N  F  F  Y  L  C  R  G  R  G  R  L  C  L  .  A  I  P  E  V  V  R  R  L  F  W  R  P  R  L  L  Q  K

CTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGG
GAGGGCCCTCGAACATATAGGTAAAAGCCTAGACTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCC 4200
 A  P  G  S  L  Y  I  H  F  R  I  .  S  R  D  R  M  R  I  V  S  H  D  .  T  R  W  I  A  R  R  F  S  G

CCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGT
GGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCA 4300
  R  L  G  G  E  A  I  R  L  .  L  G  T  T  D  N  R  L  L  .  C  R  R  V  P  A  V  S  A  G  A  P  G

TCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCT
AGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGA 4400
 S  F  C  Q  D  R  P  V  R  C  P  E  .  T  A  G  R  G  S  A  A  I  V  A  G  H  D  G  R  S  L  R  S

GTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAG
CACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTC 4500
 C  A  R  R  C  H  .  S  G  K  G  L  A  A  I  G  R  S  A  G  A  G  S  P  V  I  S  P  C  S  C  R  E  S

TATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTAC
ATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATG 4600
  I  H  H  G  .  C  N  A  A  A  A  Y  A  .  S  G  Y  L  P  I  R  P  P  S  E  T  S  H  R  A  S  T  Y

TCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCC
AGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGG 4700
 S  D  G  S  R  S  C  R  S  G  .  S  G  R  R  A  S  G  A  R  A  S  R  T  V  R  Q  A  Q  G  A  H  A

GACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGG
CTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACC 4800
 R  R  R  G  S  R  R  D  P  W  R  C  L  L  A  E  Y  H  G  G  K  W  P  L  F  W  I  H  R  L  W  P  A  G

GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTAT
CACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATA 4900
  C  G  G  P  L  S  G  H  S  V  G  Y  P  .  Y  C  .  R  A  W  R  R  M  G  .  P  L  P  R  A  L  R  Y

CGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCA
GCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGT 5000
 R  R  S  R  F  A  A  H  R  L  L  S  P  S  .  R  V  L  L  S  G  T  L  G  F  E  M  T  D  Q  A  T  P
```

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence Page 1

```
ACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGG
TGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCC 5100
N L P S R D F D S T A A F Y E R L G F G I V F R D A G W M I L Q R G

GGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA
CCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGT 5200
D L M L E F F A H P N L F I A A Y N G Y K . S N S I T N F T N K A

TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
AAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTA 5300
F F S L H S S C G L S K L I N V S Y H V C I P S T S S . S L A . S

GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGT
CCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACGGATTACTCA 5400
V S . L F P V . N C Y P L T I P H N I R A G S I K C K A V G A . . V

GAGCTAACTCACATTAATTGCGTTG
CTCGATTGAGTGTAATTAACGCAAC 5425
S . L T L I A L
```

Tuesday, 18 November 1997 13:56 Page 1
fig 53 pLM6 (1 > 4947) Site and Sequence
Enzymes : All 146 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

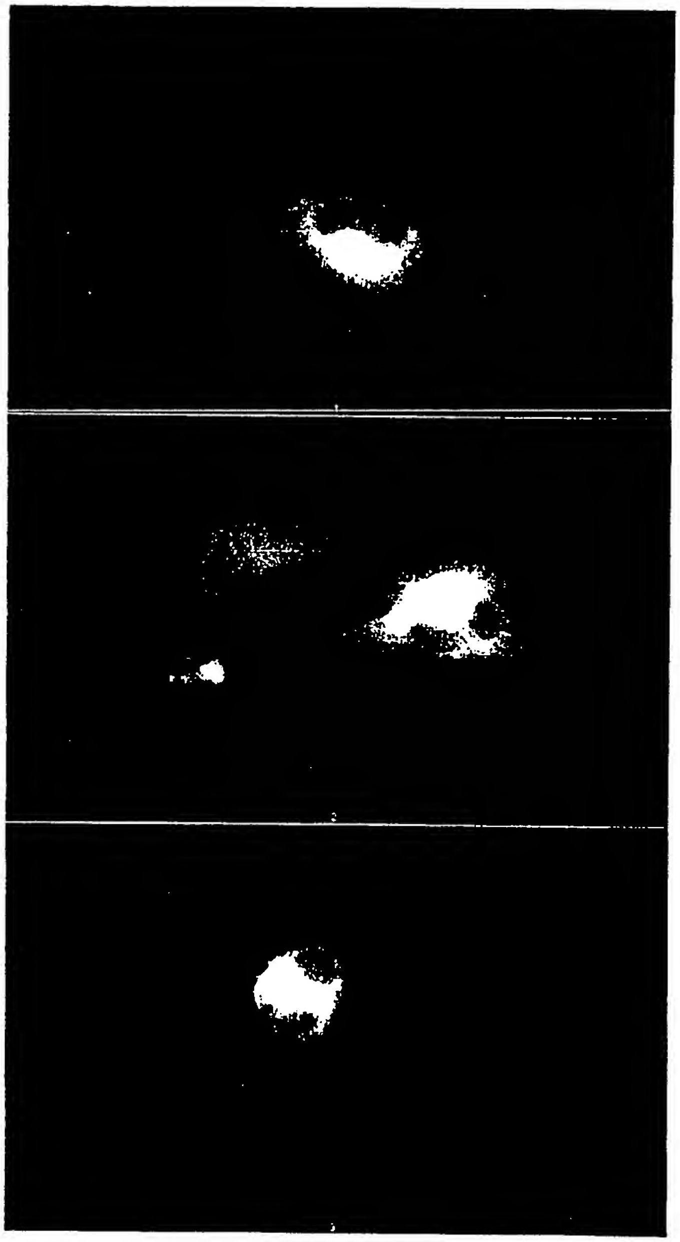
```
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
CACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTA 100
 V  H  F  S  G  K  C  A  R  N  P  Y  L  F  I  F  L  N  T  F  K  Y  V  S  A  H  E  T  I  T  L  I  N

GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTG 200
 A  S  I  I  L  K  K  E  E  Y  E  Y  S  T  F  P  C  R  P  Y  S  L  F  C  G  I  L  P  S  C  F  C  S

CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
GGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAA 300
 P  R  N  A  G  E  S  K  R  C  .  R  S  V  G  C  T  S  G  L  H  R  T  G  S  Q  Q  R  .  D  P  .  E  F

TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
AAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTCGTTGAGCCAGC 400
 S  P  R  R  T  F  S  N  D  E  H  F  .  S  S  A  M  W  R  G  I  I  P  Y  .  R  R  A  R  A  T  R  S

CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
GGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGG 500
 P  H  T  L  F  S  E  .  L  G  .  V  L  T  S  H  R  K  A  S  Y  G  W  H  D  S  K  R  I  M  Q  C  C

ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA
TATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATT 600
 H  N  H  E  .  .  H  C  G  Q  L  T  S  D  N  D  R  R  T  E  G  A  N  R  F  F  A  Q  H  G  G  S  C  N

CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT
GAGCGGAACTAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGA 700
 S  P  .  S  L  G  T  G  A  E  .  S  H  T  K  R  R  A  .  H  H  D  A  C  S  N  G  N  N  V  A  Q  T

ATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCG
TAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGC 800
 I  N  W  R  T  T  Y  S  S  F  P  A  T  I  N  R  L  D  G  G  G  .  S  C  R  T  T  S  A  L  G  P  S

GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG
CGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATC 900
 G  W  L  V  Y  C  .  .  I  W  S  R  .  A  V  V  S  R  Y  H  C  S  T  G  A  R  W  .  A  L  P  Y  R  S

TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCA
AATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGT 1000
 Y  L  H  D  G  E  S  G  N  Y  G  .  T  K  .  T  D  R  .  D  R  C  L  T  D  .  A  L  V  T  V  R  P

AGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCT
TCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGA 1100
 S  L  L  I  Y  T  L  D  .  F  K  T  S  F  L  I  .  K  D  L  G  E  D  P  F  .  .  S  H  D  Q  N  P

TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA
ATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTT 1200
 L  T  .  V  F  V  P  L  S  V  R  P  R  R  K  D  Q  R  I  F  L  R  S  F  F  S  A  R  N  L  L  L  A  N
```

Tuesday, 18 November 1997 13:56

fig 53 pLM6 (1 > 4947) Site and Sequence

Page 2

```
CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAA
GTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTT 1300
 K  K  T  T  A  T  S  G  G  L  F  A  G  S  R  A  T  N  S  F  S  E  G  N  W  L  Q  Q  S  A  D  T  I

TACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
ATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGA 1400
 Y  C  P  S  S  V  A  V  V  R  P  P  L  Q  E  L  C  S  T  A  Y  I  P  R  S  A  N  P  V  T  S  G  C

GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC
CGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCG 1500
C  Q  W  R  .  V  V  S  Y  R  V  G  L  K  T  I  V  T  G  .  G  A  A  V  G  L  N  G  G  F  V  H  T  A

CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGG 1600
 Q  L  G  A  N  D  L  H  R  T  E  I  P  T  A  .  A  M  R  K  R  H  A  S  R  R  E  K  G  G  Q  V  S

GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTT
CCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAA 1700
 G  K  R  Q  G  R  N  R  R  A  H  E  G  A  S  R  G  K  R  L  V  S  L  .  S  C  R  V  S  P  P  L  T

GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG
CTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAAC 1800
.  A  S  I  F  V  M  L  V  R  G  A  E  P  M  E  K  R  Q  Q  R  G  L  F  T  V  P  G  L  L  L  A  F  C

CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCG
GAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGC 1900
 S  H  V  L  S  C  V  I  P  .  F  C  G  .  P  Y  Y  R  L  .  V  S  .  Y  R  S  P  Q  P  N  D  R  A

CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTT
GTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAA 2000
 Q  R  V  S  E  R  G  S  G  R  A  P  N  T  Q  T  A  S  P  R  A  L  A  D  S  L  M  Q  L  A  R  Q  V

CCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT
GGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACA 2100
S  R  L  E  S  G  Q  .  A  Q  R  N  .  C  E  L  A  H  S  L  G  T  P  G  F  T  L  Y  A  S  G  S  Y  V

TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCT
ACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATACTGGTACTAATGCGGTTCGCGCGTTAATTGGGAGTGATTTCCCTTGTTTTCGA 2200
 V  V  N  C  E  R  I  T  I  S  H  R  K  Q  L  .  P  .  L  R  Q  A  R  N  .  P  S  L  K  G  T  K  A

GGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCATGCAAATGAGGAGGAGGAGCCAGAGA
CCCATGGCCCGGGGGGGAGCTCCAGCTGCCATAGCTATTCGAACTATAGCTTAAGGACGTCGGGCCCCCTAGGTACGTTTACTCCTCCTCCTCGGTCTCT 2300
                                                                                   U3 stuk
 G  Y  R  A  P  P  R  G  R  R  Y  R  .  A  .  Y  R  I  P  A  A  R  G  I  H  A  N  E  E  E  E  P  E
```



a

b

c

Figure 43

donderdag, 27 november 1997 16:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence
Enzymes : All 146 enzymes (No Filter)
Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAGAAATTGGAGCAACTACC 100

CACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAATTTTGAGTTTACGACTACAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCTATT 200

CTGTATGACTAGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAAAGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTTAGTTTTTCGATC 300

AACTAGATCTTACAAAACTTGCAATACAATTCCATTTTCAGATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCAGCAACTAACCCAAATTCCA 400

ACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGTATAAAATGCC 500

CAGAAAAGAAGATGATAAAAATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTCTAATATGACGGCTTATATCTCAATTTTCTTGAGTTTTATCAAA 600

AAATTTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAATTTTGTCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAA 700

AGTGGAAGGAGCGCAAGTCTATACTGGAAATAATGATCTGAAACAAATTTGTGCTATTCTCAAATGTTTAAGACATGTTTTGAAGATTTTTTCAAATTCG 800

CACTAGTTTCAGAACCTTCCTTTTTGTATGAAAAAGTAAAAAAAAACTATTTCAAACCCTCACCGCCACCATGTTTCAACTCTTAATTTTTATAAAATTT 900

TGCAATTTACAAATCGCCTCCCCTTGCCCGAAAAGTGCCCACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATTGATGTGAGAAAACACAGAAG 1000

AGGCTAACTAAATTGACAGGGACAGGTTGTCCCTCTTCTCCCTCCTTCTCCCGCCTCCTCCTCCCGTTCCATCTCCAACAACAACAATTTTCCAATTTCG 1100

TTGTCCATTTTGCTTATAAACATTTGTGTGTGGAAGGAAACTACACGGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACTCTTTCGGAAAT 1200

TGATGAATAAAGATAGAGCCGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTTTTTCATCGTCTCAACTTGCGCATGAGTCTTCCCCCGCTCT 1300

CATCACTGACAATTAATGTCGGGTTTTATGCGCTCTTTCCTATTCCGCCACTCATTCTGGGTTACCACAAACTGGAATACATTTTACTACTATTCAAGCC 1400

ATTTATTTTGATATTTAATTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACGATATTTTCGATTTTAAAAAATCTGAAAAGT 1500

TTCAAAAAATCAATAAATATTCCCTAACAAATTGTATGGCTAAAATTTTATTTCTACTGTTGACAATATCTTTATATGTATCACTGTTTTCCATCTCAAA 1600

ACCTTGAATCCCCCAAGTTATAGGAAGCTCCGTGTCACATTTCCCATGCTATGAATCGCTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGATAA 1700

TTATTGGGCACGCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAAGCACAGATTATCTATCAAATTCAATCTTTGAACATTCAGCCAGTTCGTACAA 1800

TTTTCCATGCTTTTTGGCCCATTAAAAAACTTTCTCACCTCTTCATCCATCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACTCTACTTTTTAAG 1900

AGAAGGAGATACTGAGCCACATGGCGTGTGACCCTTTTCATCTCGTCCGTTCGGTCTCAAATTCACGCTCATACTAACTCTTCAAATAGCCATAGACCTC 2000

CTTGTTTTCTTCTTCGTTTTGACTCGCGCCTATTTTTTTGTGGCTGCCTGAAAGCCGGGAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100

AAAAACGAGGCAATTTAAAAATATTAAAATTAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAAACAAAACAAATAGGAACCGCCAGATCAAAA 2200

TTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGGCTGAACAGAAAACTGAAGTGCCTGCATAAATCTAGTGTAACGTTTAGATTGAACTCGGAA 2300

ATCCTAAGCCTGAACTATAGCCTTATTCTAGATCTTAGTTGCGCATAAGCTCAAGCCCAAGCAGAAATGACTTGCATTTAGTTTAAGCCTAGATTGACTT 2400

GCTTGCTTCAGTCTAATCCAGACTAGATTTCCAAGAGAGTTTTCAATTTTAAATGTTTCCAGTTTCTTGTTACTTAAAATCTTAATGCCCTGTGATGCGT 2500

AAAATCGTTATCCCTTTCTCTCACACTTTCAATTACAGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACC 2600

ACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAATGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTA 2700

TCCGATCCTTCCGGCTTCTTTTTAGAAATTATATTATTTCAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800

AAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAA 2900

CTTGCTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACC 3000

donderdag, 27 november 1997 16:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence Page 2

CATCTTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCT 3100
GAAGCCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAA 3200
CAAGACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAA 3300
GCATGCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACA 3400
GCCGATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAA 3500
GATCCACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATAT 3600
CTGAAAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGAC 3700
TTCAATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAA 3800
GACTCCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAG 3900
AATCCAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGG 4000
CTATCCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGAC 4100
ATGGCAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTCGGTCCTCCACAT 4200
CAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGG 4300
ACAACATTCGCTAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTATGCACTCACAGACTAGTCGA 4400
CGACCTTCTTCACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATCCACCGAGCACAGAATGGCGG 4500
CTCTCTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGCATCCTACTCGGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTA 4600
TGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGCTAGCACGACA 4700
GCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGTCGACTCACTAACCAAGAAAC 4800
AGGAGAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACG 4900
ATTCAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCTTCGTCAACCA 5000
TCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGCTCGTTTGGCAAGAACAAGA 5100
AGAGCTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAACTCT 5200
TGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTT 5300
GATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTT 5400
CCCGCGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTG 5500
CAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCA 5600
ACCAGTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGCAGCATCAACTTGGAATCGATG 5700
CTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCATCCAACTGACATTCTTACTTC 5800
CTCAACTACAATCCGAATGTTCATGCACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACTC 5900
GTCAAGTCAATTTTGACAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTATGTATCTATTC 6000

dondordag. 27 november 1997 16:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence Page 3

GAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAG 6100
AAGCAAAGAATCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGTCCCACTTCAAAACAACGAA 6200
GGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAAATGTCAGTAATGTCGAATCGTCTCGAAGGAT 6300
TCATCCTACGTTACCTCCGACGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAAT 6400
AGCTCTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACTGTCCTCTAACTGTCGATGGA 6500
TCCCGTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCACTTCCT 6600
TCGAGGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAAGACCTCGTCCC 6700
GTCACCTGCCAACTCATCCCGACAACACTTCAATCCCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAA 6800
GACTCTAATCTTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCATGGTATTGATATCTGAGCTCCGCATCGGCCGCTGTCATCAGATCG 6900
CCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTAT 7000
CAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAA 7100
AAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCT 7200
GATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACG 7300
TCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCC 7400
TGCTTTTGCTTTTTGGGGGCTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGA 7500
AAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGA 7600
CAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCT 7700
GACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCG 7800
GGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATC 7900
AGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTG 8000
CGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA 8100
AGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAA 8200
AGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCA 8300
ATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATG 8400
ACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGC 8500
GGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCG 8600
GAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC 8700
TAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA 8800
ATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAG 8900
GCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGA 9000

donderdag, 27 november 1997 16:48 Page 4
fig 44 pNP9 Map (1 > 10122) Site and Sequence

TTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTG 9100

AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCG 9200

GTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGT 9300

AGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT 9400

TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTAC 9500

ACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAG 9600

GAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTC 9700

GTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTA 9800

TCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAG 9900

CGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGA 10000

GCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAA 10100

CAATTTCACACAGGAAACAGCT 10122

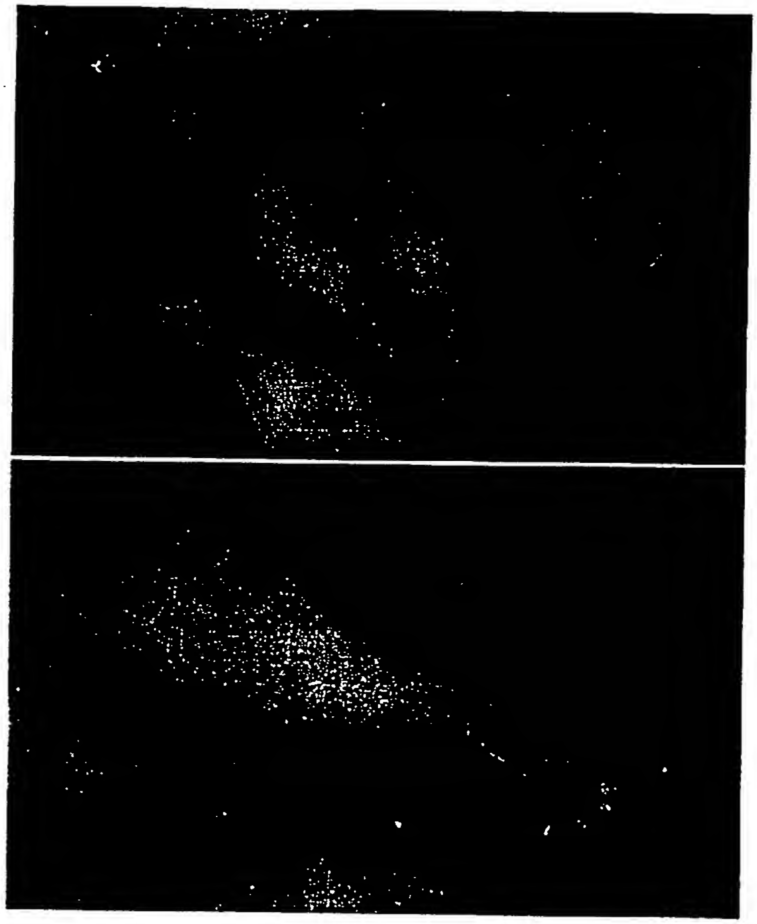
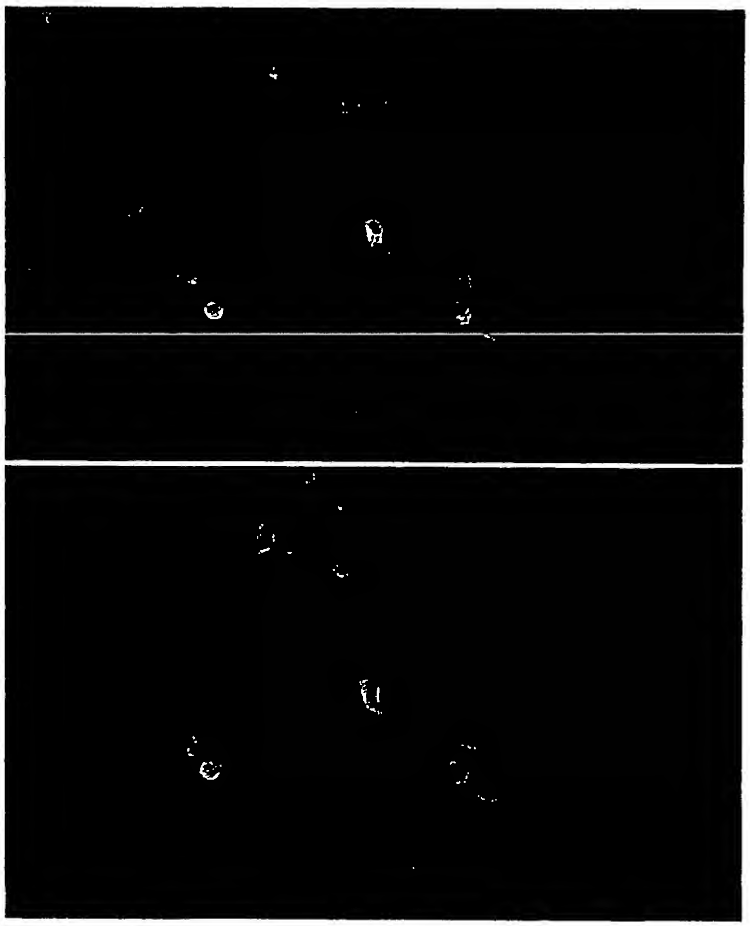

a
b

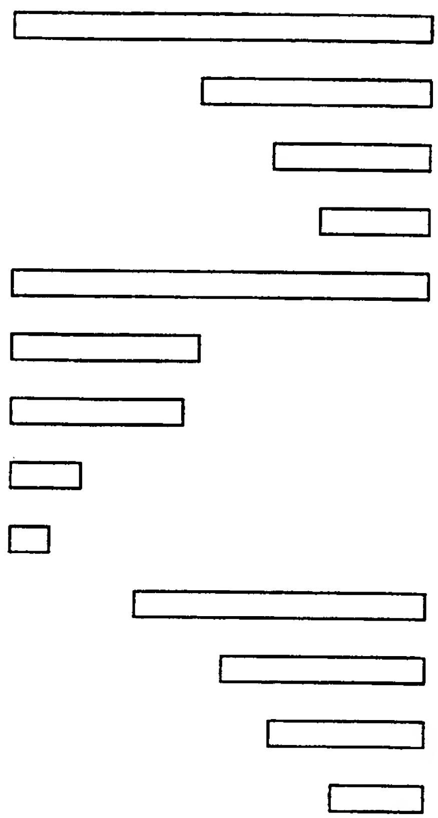
figure 45



a

b

Figure 46



	+ end	MTB	actin
pTB72	+	+	+
pcDU2	nd	nd	+
pcDU3	nd	nd	+
pcDU4	nd	nd	+
peGFP72	+	+	nd
peGFPsma	- nd	+	
peGFPecl	-	+	nd
peGFPxba	-	-	nd
peGFPsac	-	-	nd
pLM4	+	+	nd
pCB251	nd	nd	nd
pLM5	nd	nd	nd
pCB201	nd	nd	+

Figure 47

a

b

c

Figure 50

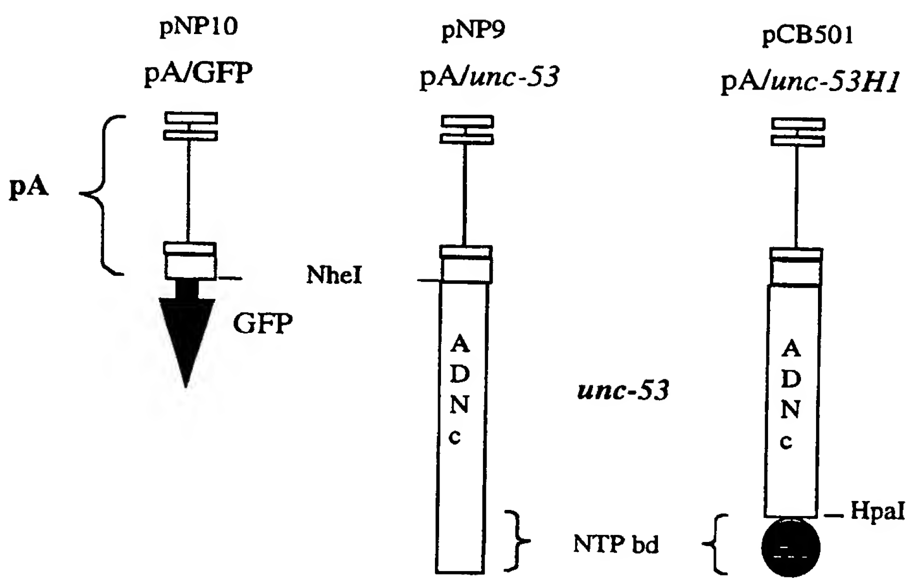

pNP10
pA/GFP
pNP9
pA/*unc-53*
pCB501
pA/*unc-53H1*
pA
NheI
GFP
A
D
N
c
unc-53
A
D
N
c
HpaI
NTP bd

Figure 51a

donderdag, 27 november 1997 16:46
fig 35 pNP9 Map (1 > 12641) Site and Sequence Page 3

GATTTTCAGACTTGGGCTATAAATTTTTGTCAAAACTAGGAATCTTAAAATATTTGTATTTTTCGAAGAATGCTCCTCAATCTCAAATTCATATTTTATA 6100
ATTTCAGACCCCGTGATATCTGAAAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAG 6200
TTGTTCCACTTAAAATGACTTCAATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGA 6300
GCAGTCGTCCGCGTCTGAAGACTCCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATG 6400
GGAAAGAATAAGACATCAGAATCCAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGA 6500
CTCGTCGAGCACAGAACGGCTATCCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGA 6600
CGATTTGTCCGGAGTAGACATGGCAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTC 6700
CCCAGTCGGTCCTCCACATCAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCAACGTGGCGCCG 6800
CTGCCACCTCAACCTTCGGACAACATTCGCTAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTAT 6900
GCACTCACAGACTAGTCGACGACCTTCTTCACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATCC 7000
ACCGAGCACAGAATGGCGGCTCTCTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCCTACTCGGCGCGTTCCCGAGGTG 7100
GAAGCTCTACTGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACT 7200
ATCACTGGCTAGCACGACAGCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGTC 7300
GACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTTGA 7400
AGCCTGAAGAGGCAATACGATTCAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGG 7500
TGAGCTTCTTCGTCAACCATCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGC 7600
TCGTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATTT 7700
CCGGATCTCAAGGAACTCTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGCCTTGACAATCT 7800
GGATCGTGCCCGCGAAGTTGATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGT 7900
CCAGCCACTCGTGCTTCTTCCCGCGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTT 8000
CGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAGAGATAATCGT 8100
AGGATATCTTGCCATGTCAACCAGTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAG 8200
CATCAACTTGGAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACTTCCACGCGTCATTGGACACTCCACAACCATGATAACCAGCCATC 8300
CAACTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCATGCACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAA 8400
GCAAATGATTCTCCAACTCGTCAAGTCAATTTTGACAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTG 8500
GCTGCTTATGTATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAACGAC 8600
GCCTGGAAAAGATCTTGAGAAGCAAAGAATCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGT 8700
CCCACTTCAAAACAACGAAGGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAAATGTCAGTAATG 8800
TCGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGACGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAA 8900
TCATTGACTTCTTCCCAATAGCTCTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACTG 9000

donderdag, 27 november 1997 16:46
fig 35 pNP8 Map (1 > 12641) Site and Sequence Page 4

TCCTCTAACTGTCGATGGATCCCGTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACC 9100
TTCGGTCGCTGCACTTCCTTCGAGGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCTCAAACGTCTTC 9200
AACTCCAAGACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCAT 9300
CGACAACATTTGAACAGAAGACTCTAATCTTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCATGGTATTGATATCTGAGCTCCGCATC 9400
GGCCGCTGTCATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCAC 9500
CCCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAA 9600
TAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTC 9700
TTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCA 9800
ATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAG 9900
TTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTAT 10000
TGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATG 10100
GGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGT 10200
GATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAG 10300
CGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGT 10400
AAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGG 10500
CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTT 10600
CAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG 10700
AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCG 10800
CCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGC 10900
ATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAA 11000
CCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCG 11100
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTA 11200
ACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTG 11300
GCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT 11400
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTT 11500
TACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGCTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC 11600
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA 11700
AAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCACATACCAAATACT 11800
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA 11900
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAG 12000

donderdag, 27 november 1997 16:46

fig 35 pNP8 Map (1 > 12641) Site and Sequence Page 5

CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTA 12100

AGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGC 12200

GTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCA 12300

CATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGC 12400

GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCG 12500

ACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTG 12600

TGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCT 12641

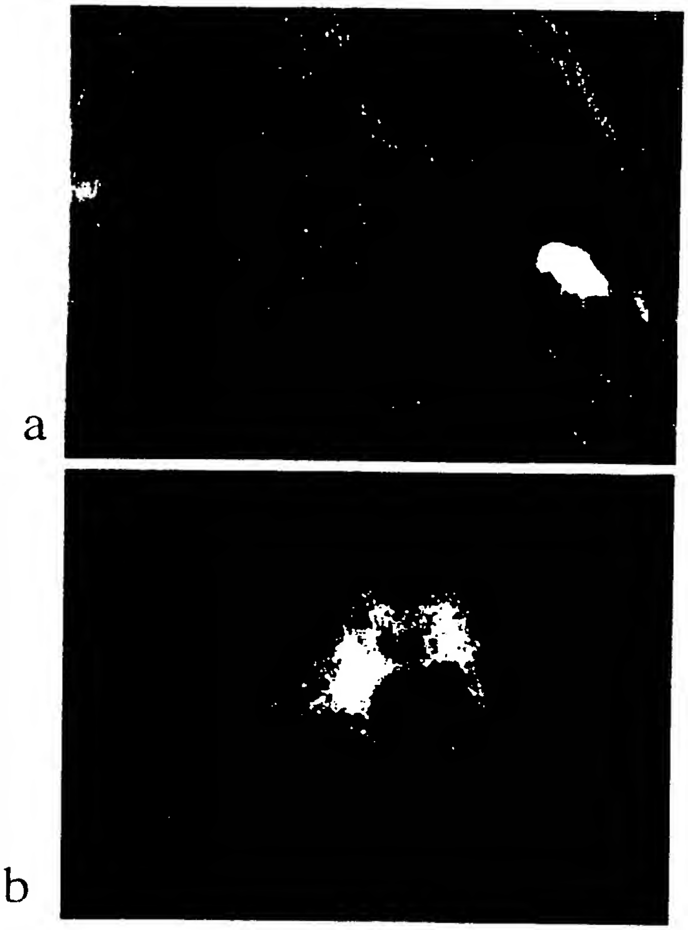
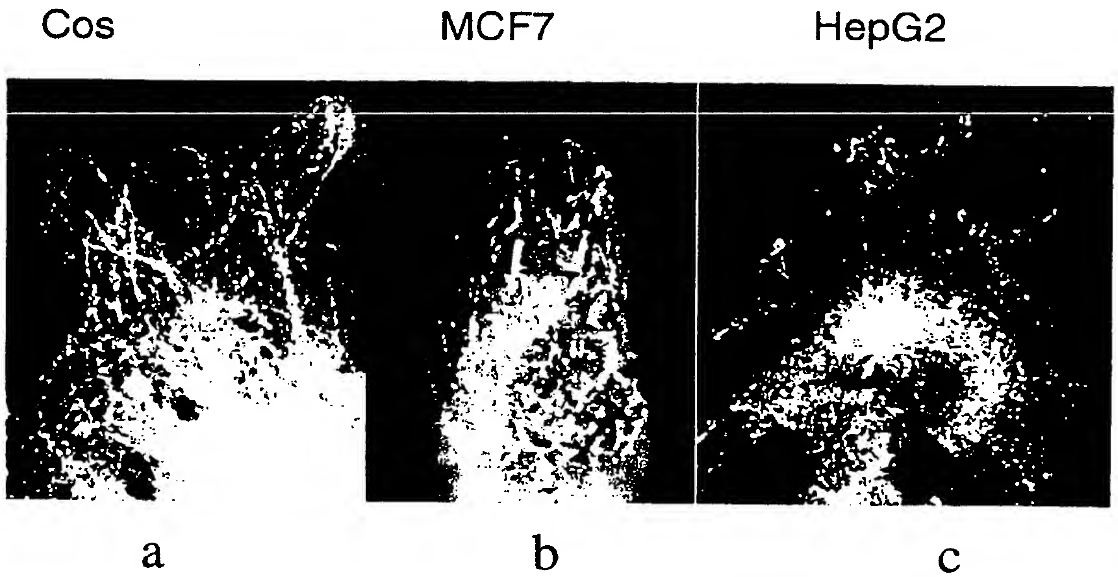
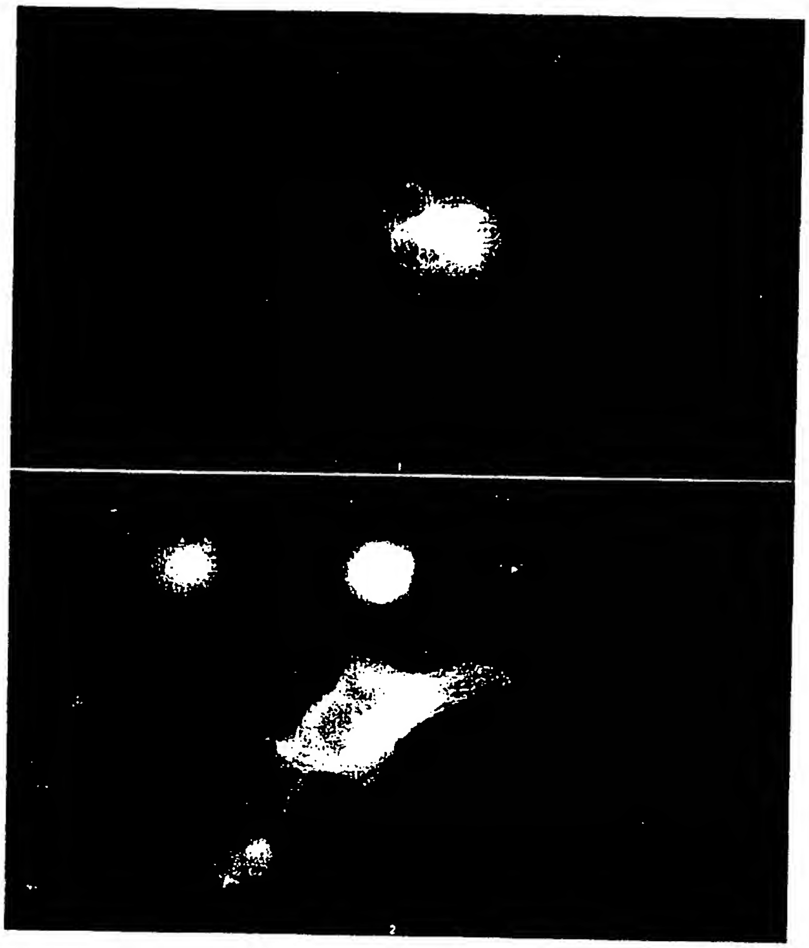


Figure 36: Accosiation of *C. elegans* UNC-53 (expressed from pTB72) with the microtubular cytoskeleton of HepG2 cells. (A) microtubules stained with YL1/2 antibody to tubulin and (B) Staining for *C. elegans* UNC-53.



Fgiure 37: Microtubule (+)-end binding of *C. elegans* UNC-53 following traneint transfection with pTB72 of HepG2 (a), MCF7 (b) and Cos cells (c). *C. elegans* UNC-53 was visualised by immunofluorescence using mab16-48.



a

b

Figure 38

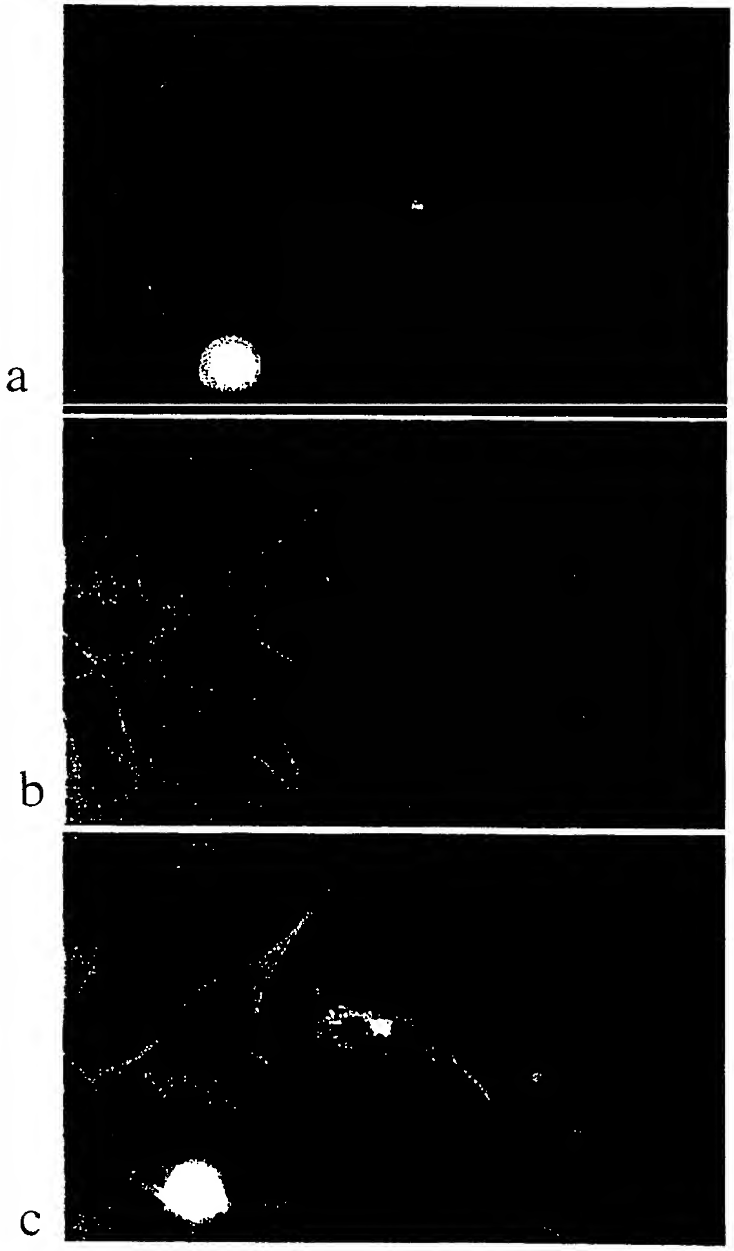


Figure 39

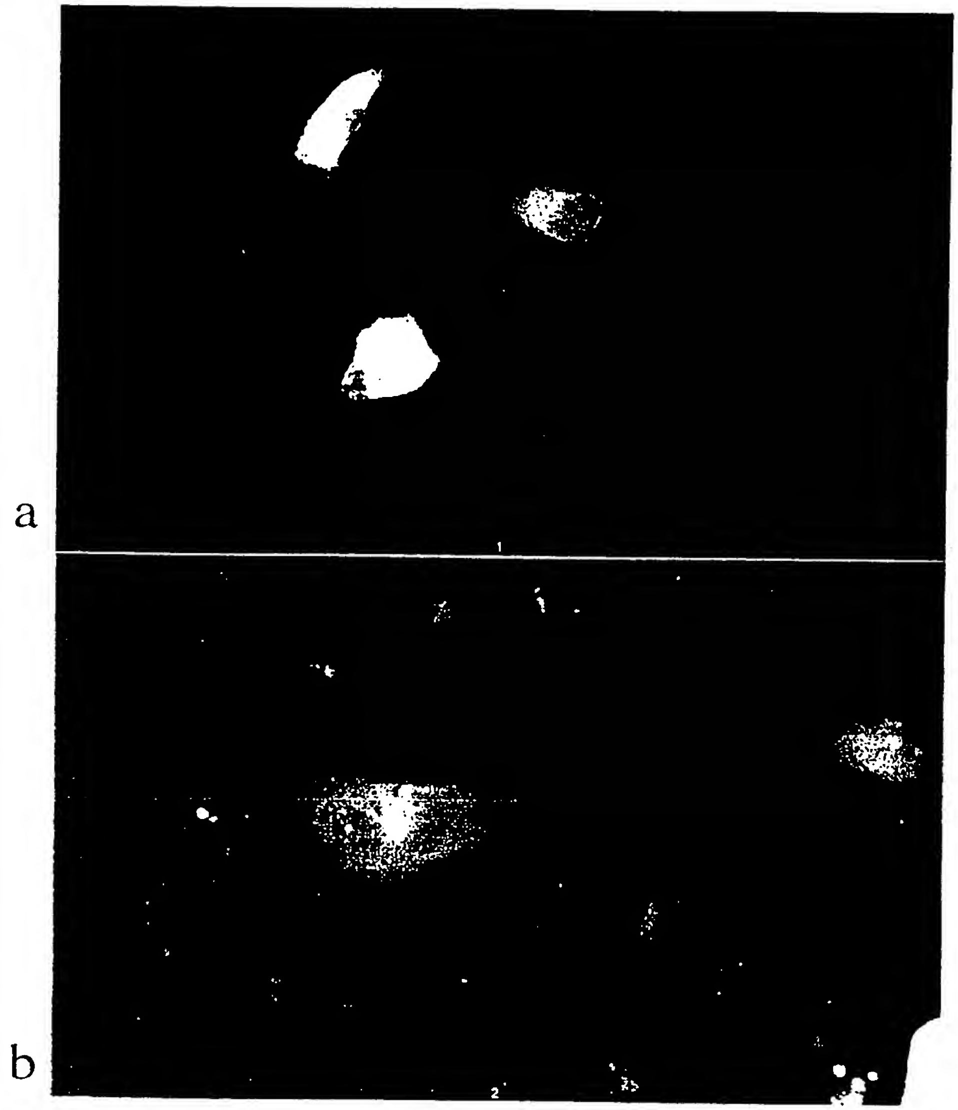
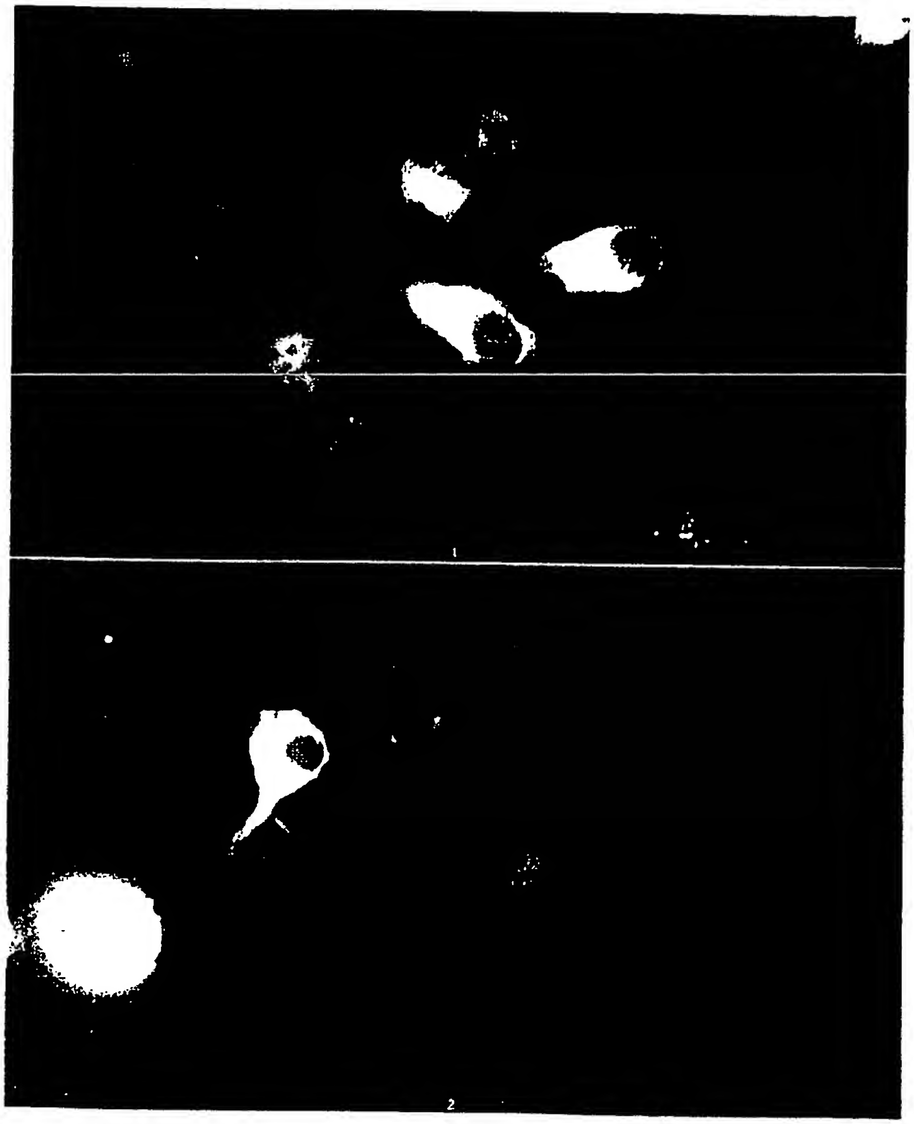


Figure 40

a



b

Figure 41

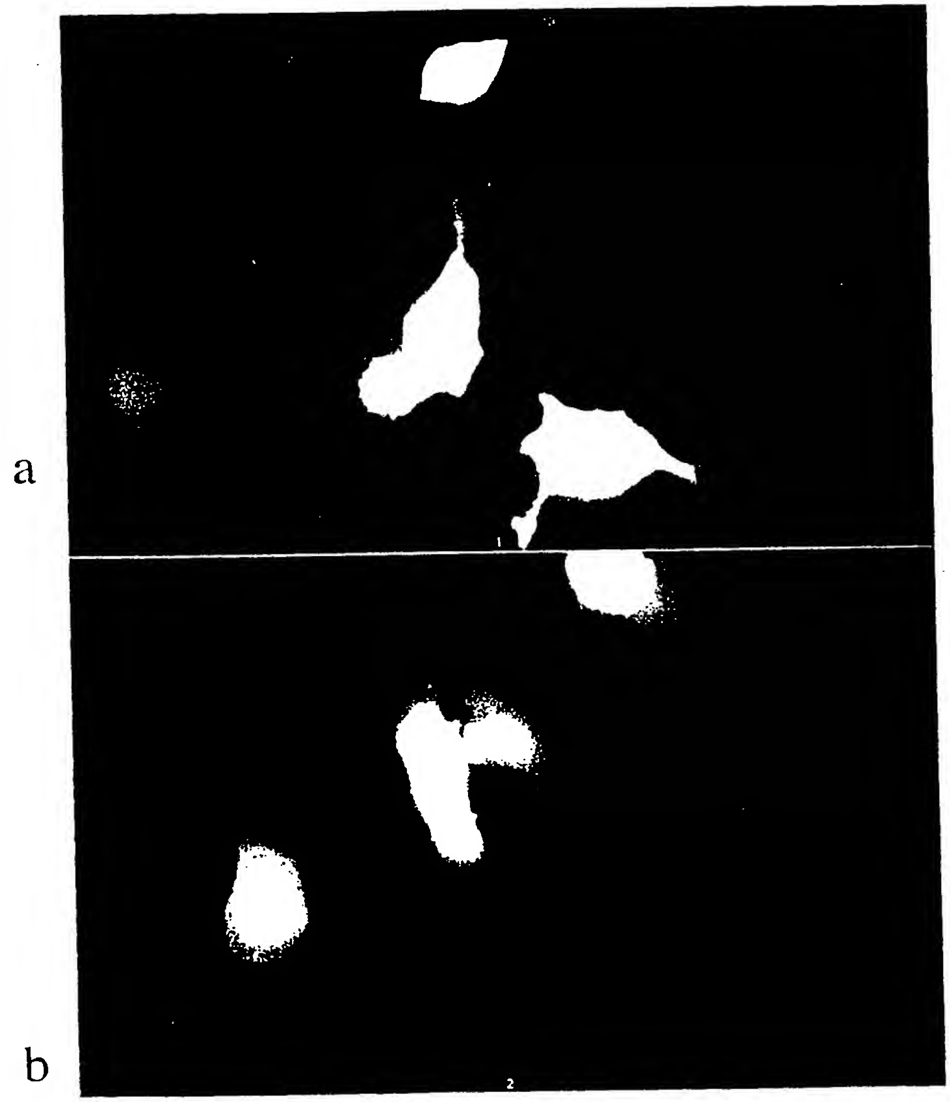


figure 42

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 7

CATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGA
GTATCTGAAAGACTTCTTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCT 4200

insert pLM1

ORF pLM1

I D F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R

CAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGA
GTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCT 4300

insert pLM1

ORF pLM1

Q N S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K

GTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTAC
CAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATG 4400

insert pLM1

ORF pLM1

S V V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T

ACCCGACTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCCACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACT
TGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGA 4500

insert pLM1

ORF pLM1

P D S S A P S S P K L Q H G S T E T A S P S I K S S T L S S V G T

GATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCT
CTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGA 4600

insert pLM1

ORF pLM1

D V T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R

CTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACAT
GACTCGATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTA 4700

insert pLM1

ORF pLM1

S E L W E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

GCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCT
CGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGA 4800

insert pLM1

ORF pLM1

Q L E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S

GCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTA
CGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCAT 4900

insert pLM1

ORF pLM1

A L S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S

CTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGG
GAACACCAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCC 5000

insert pLM1

ORF pLM1

T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G

CTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACC
GACATCGTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGG 5100

insert pLM1

ORF pLM1

C S K V S G K V D W K M L D E A V F Q V F K D Y I S K M D P A S T

CTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTG
GACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCAC 5200

insert pLM1

ORF pLM1

L G L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G

TCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACAT
AGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTA 5300

insert pLM1

ORF pLM1

V N N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 9

AAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGC
TTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCG 5400

insert pLM1

ORF pLM1

S L L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R

TCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGA
AGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCT 5500

insert pLM1

ORF pLM1

S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q

TAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCAC
ATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTG 5600

insert pLM1

ORF pLM1

I D R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T

CTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACC
GACGTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGG 5700

insert pLM1

ORF pLM1

C K Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T

TTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGC
AAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCG 5800

insert pLM1

ORF pLM1

F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E

TGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTT
ACGAAGCCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAA 5900

insert pLM1

ORF pLM1

L L R V L D W V P K L W Y H L H T F L E K H S T S D F L I G P C F F

TCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGG
AGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCC 6000

insert pLM1

ORF pLM1

L S C P I G I E D F R T W F I D L W N N S I I P Y L Q E G A K D G

ATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGT
TATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACA 6100

insert pLM1

ORF pLM1

I K V H G Q K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L

ACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCC
TGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGG 6200

insert pLM1

ORF pLM1

Y H L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S D P

TCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAA
AGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATT 6300

insert pLM1

ORF pLM1

L M A M L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L .

GGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAG
CCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTC 6400

insert pLM1

G F G N H C H P R T A E R W H Q L S . L L L S P L L F Q S T G S P

CCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGC
GGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCG 6500

insert pLM1

A P G G E Q E G G G D E R G G T G S W C C T F E N F L G R N G G V A

GTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATT
CAAACCCTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAA 6600

insert pLM1

F G N L C P L N T F T G L L . . L W G K D D S G S F P . L L V S I

Tuesday, 18 November 1997 11:48 Page 11
fig 34 pLM4 (1 > 10070) Site and Sequence

```
ACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGGCTGAACTTGCTCA
TGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAGTTTTGTGACGTCGTCAAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGT 6700
insert pLM1
T N S W A F W G G V Q K T S K H C S S S P E F S L D L T R L N L L

AAAGAAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTTCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA
TTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAATGATCAAGATCTATTGACTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTT 6800
insert pLM1
K R S R I P A H W R P L L V L D N . S . S A I P H L . R F Y L L .

AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATA
TTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTAT 6900
T S H T S P . T . N I K . M Q L L L L T C L L Q L I M V T N K A I

GCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTA
CGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAACATTCGCAAT 7000
f1 ori
A S Q I S Q I K H F F H C I L V V V C P N S S M Y L N A . I V S V

ATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGAC
TATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTG 7100
f1 ori
N I L L K F A L N F C . I S S F F N Q . A E I G K I P Y K S K E . T

CGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
GCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTA 7200
f1 ori
E I G L S V V P V W N K S P L L K N V D S N V K G R K T V Y Q G D

GGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTT
CCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAA 7300
f1 ori
G P L R E P S P . S S F L G S R C R K A L N R N P K G S P R F R A

GACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAAC
CTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTG 7400
f1 ori
. R G K P A N V A R K E G K K A K G A G A R A L A S V A V T L R V T

CACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATA
GTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTAT 7500
f1 ori
T T P A A L N A P L Q G A S G G T F R G N V R G T P I C L F F . I
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

CATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAA
GTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAGGACTCCGCCTTTCTTGGTCGACACCTT 7600
H S N M Y P L M R Q . P . . M L Q . Y . K R K S P E A E R T S C G

TGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCC
ACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGG 7700
M C V S . G V E S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P

CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGC
GGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCG 7800
Q A P Q Q A E V C K A C I S I S Q Q P . S R P . L R P S R P . L R

CCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGG
GGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCC 7900
P V P P I L R P M A D . F F L F M Q R P R P P R P L S Y S R S S E

AGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
TCCGAAAAAACCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGG 8000
Kan/Neo
E A F L E A . A F A K I D Q E T G . G S F R M I E Q D G L H A G S P

GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCG
CCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGC 8100
Kan/Neo
A A W V E R L F G Y D W A Q Q T I G C S D A A V F R L S A Q G R P

GTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAG
CAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTC 8200
Kan/Neo
V L F V K T D L S G A L N E L Q D E A A R L S W L A T T G V P C A

CTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAA
GACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTT 8300
Kan/Neo
A V L D V V T E A G R D W L L L G E V P G Q D L L S S H L A P A E K

AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGT
TCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCA 8400
Kan/Neo
V S I M A D A M R R L H T L D P A T C P F D H Q A K H R I E R A R

ACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGC
TGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCTCGTACG 8500
Kan/Neo
T R M E A G L V D Q D D L D E E H Q G L A P A E L F A R L K A S M

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence


```
CCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGC
GGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGA 8600
Kan/Neo
P D G E D L V V T H G D A C L P N I M V E N G R F S G F I D C G R L

GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT
CCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCA 8700
Kan/Neo
G V A D R Y Q D I A L A T R D I A E E L G G E W A D R F L V L Y G

ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC
TAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGG 8800
Kan/Neo
I A A P D S Q R I A F Y R L L D E F F . A G L V G S K . P T K R R

CAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGC
GTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCG 8900
P T C H H E I S I P P P P S M K G W A S E S F S G T P A G . S S S A

GGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAA
CCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATGGCCTTCCTTGGGCGCGATACTGCCGTTATT 9000
G I S C W S S S P T L G G G . L K H G R R Q Y R K E P A L . R Q .

AAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTG
TTTCTGTCTTATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGGTGGCTCTGGGGTAAC 9100
K D R I K R T V L G R L F I N A G F G P R A G T L S I P H R D P I

GGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCC
CCCGGTTATGCGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTCGGTTGCAGCCCCGCCGTCCGGGACGG 9200
G A N T P A F L P F P H P T P Q V R V K A Q G S Q P T S G R Q A L P

ATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCA
TATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGT 9300
. P Q V T H I Y F R L I . N F I F N L K G S R . R S F L I I S . P

AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG
TTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGAC 9400
pUC ori
K S L N V S F R S T E R Q T P . K R S K D L L E I L F F C A . S A

CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
GAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTC 9500
pUC ori
A C K Q K N H R Y Q R W F V C R I K S Y Q L F F R R . L A S A E R R
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

ATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTC 9600
pUC ori
Y Q I L S F . C S R S . A T T S R T L . H R L H T S L C . S C Y Q

TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
ACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCAC 9700
pUC ori
W L L P V A I S R V L P G W T Q D D S Y R I R R S G R A E R G V R

CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
GTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTG 9800
pUC ori
A H S P A W S E R P T P N . D T Y S V S Y E K A P R F P K G E R R T

AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC
TCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG 9900
pUC ori
G I R . A A G S E Q E S A R G S F Q G E T P G I F I V L S G F A T

TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
AGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGAC 10000
pUC ori
S D L S V D F C D A R Q G G G A Y G K T P A T R P F Y G S W P F A

GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 10070
G L L L T C S F L R Y P L I L W I T V L P P C I

dondердag, 27 november 1997 16:46 Page 1

fig 35 pNPB Map (1 > 12641) Site and Sequence

Enzymes : All 146 enzymes (No Filter)

Settings : Circular, Certain Sites Only, Standard Genetic Code

ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAGAAATTGGAGCAACTACC 100

CACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAATTTTGAGTTTACGACTACAAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCTATT 200

CTGTATGACTAGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAAAGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTTAGTTTTTCGATC 300

AACTAGATCTTACAAAACTTGCAATACAATTCCATTTTCAGATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCAGCAACTAACCCAAATTCCA 400

ACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGTATAAAATGCC 500

CAGAAAAGAAGATGATAAAAATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTCTAATATGACGGCTTATATCTCAATTTTCTTGAGTTTTATCAAA 600

AAATTTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAATTTTGTCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAA 700

AGTGGAAGGAGCGCAAGTCTATACTGGAAATAATGATCTGAAACAAATTTGTGCTATTCTCAAATGTTTAAGACATGTTTTGAAGATTTTTTCAAATTCG 800

CACTAGTTTCAGAACCTTCCTTTTTGTATGAAAAAGTAAAAAAAAACTATTTCAAACCCTCACCGCCACCATGTTTCAACTCTTAATTTTTATAAAATTT 900

TGCAATTTACAAATCGCCTCCCCTTGCCCGAAAAGTGCCCACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATTGATGTGAGAAAACACAGAAG 1000

AGGCTAACTAAATTGACAGGGACAGGTTGTCCCTCTTCTCCCTCCTTCTCCCGCCTCCTCCTCCCGTTCCATCTCCAACAACAACAATTTTCCAATTTCG 1100

TTGTCCATTTTGCTTATAAACATTTGTGTGTGGAAGGAAACTACACGGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACTCTTTCGGAAAT 1200

TGATGAATAAAGATAGAGCCGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTTTTTCATCGTCTCAACTTGCGCATGAGTCTTCCCCCGCTCT 1300

CATCACTGACAATTAATGTCGGGTTTTATGCGCTCTTTCCTATTCCGCCACTCATTCTGGGTTACCACAAACTGGAATACATTTTACTACTATTCAAGCC 1400

ATTTATTTTGATATTTAATTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACGATATTTTCGATTTTAAAAAATCTGAAAAGT 1500

TTCAAAAAATCAATAAATATTCCCTAACAAATTGTATGGCTAAAATTTTATTTCTACTGTTGACAATATCTTTATATGTATCACTGTTTTCCATCTCAAA 1600

ACCTTGAATCCCCCAAGTTATAGGAAGCTCCGTGTCACATTTCCCATGCTATGAATCGCTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGATAA 1700

TTATTGGGCACGCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAAGCACAGATTATCTATCAAATTCAATCTTTGAACATTCAGCCAGTTCGTACAA 1800

TTTTCCATGCTTTTTGGCCCATTAAAAAACTTTCTCACCTCTTCATCCATCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACTCTACTTTTTAAG 1900

AGAAGGAGATACTGAGCCACATGGCGTGTGACCCTTTTCATCTCGTCCGTTCGGTCTCAAATTCACGCTCATACTAACTCTTCAAATAGCCATAGACCTC 2000

CTTGTTTTCTTCTTCGTTTTGACTCGCGCCTATTTTTTGTGGCTGCCTGAAAGCCGGGAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100

AAAAACGAGGCAATTTAAAAATATTAAAATTAATGAGGTTGTAGATGTAGATTTGCAAAAGAAGAAAAAAACAAAACAAATAGGAACCGCCAGATCAAAA 2200

TTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGGCTGAACAGAAAACTGAAGTGCCTGCATAAATCTAGTGTAACGTTTAGATTGAACTCGGAA 2300

ATCCTAAGCCTGAACTATAGCCTTATTCTAGATCTTAGTTGCGCATAAGCTCAAGCCCAAGCAGAAATGACTTGCATTTAGTTTAAGCCTAGATTGACTT 2400

GCTTGCTTCAGTCTAATCCAGACTAGATTTCCAAGAGAGTTTTCAATTTTAAATGTTTCCAGTTTCTTGTTACTTAAAATCTTAATGCCCTGTGATGCGT 2500

AAAATCGTTATCCCTTTCTCTCACACTTTCAATTACAGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACC 2600

ACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAATGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTA 2700

TCCGATCCTTCCGGCTTCTTTTTAGAAATTATATTATTTCAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800

AAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAA 2900

CTTGCTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACC 3000

CATCTTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCT 3100
GAAGCCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGGTGAGTATTTGAAATCGGAAATTGGAAATGTATTTTTTAAAAACT 3200
GAAAATTCTACAAAATAAATTAAAAATAAGATTTTTTCCTCTGATAGTATTGCATCCCACTATTTTACTTTGAAGATTTATATCTTGGTTCATATTGAAG 3300
ATATCAGATATAGAAAAAGAAATAAAAAATATTTTGACAGTTGATAATTTTTGTATAGGACCAAAGACAAGTGAGATATAAGCTGTCAAAGTTGATTTTC 3400
AAGAAATTTTAAAACCCTAGTTTTGCGAAGCTCTGGGCCTCATCTATATTTAGAACCCGATTCGTACTTCTTCCGTTCCTTGACTCTACCAAAACCAAAA 3500
CCAACCTACTAATAAAAATGATGAGACAATTGGGAATTGTCTCCCATTTTCTTCTTTCTCCTTGACACTCTTTCAGAATCTATGTCCCATCTTTTTGTCG 3600
TTGTGTCCCCCCATAAAGACTCTTCGCGGAAAAATGTTGCAACGGAAGTGATATTCCGAGCATTTTTCGACGTCGAGGGCCGAAAAACACATCTGGCTGA 3700
CAAAGAGTAAAGCAATTTCTCAGCTCTTCTTCGCCGGTTTTTCAATTCGTTTTTCAAAATGAGCTACTACAGAGTGAAAGAGCACAAATTGCAAAACATT 3800
TTTGTGTGAGATGCACTTTTTGAAAAATTAACTTTACGTTTTCAGTTTCTAGTATTTATTTTTTTCATATAAATTAGAGCTTCTTAGACCTGCTATATTT 3900
TTTAAAACTTCCTACTGAAATATACGAGATTCTTTTGACTTTCCGGAATTGTCTTATGGCTTCTATTATTTATGAGAAAACATTTTTTAAAAATTTTTTT 4000
GAAAAAAAAACTGTGCATCTTCGTTTTTTACATAGTAATTTCCAGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAGACTCGA 4100
AACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCATGCATTC 4200
CACATCTTCCAAGGTTCGTTGTTTAGGAGAACTGTTTTTTTTTGTTTTCGTGACCTTCACATAGTCTCGGATGTTTATAAAAGTGAGGTCTCTGGGACAC 4300
CTGCCATAAAATGTGAATCCGCCCATTTGTTGGTACAAAAAACTTTGCAGAGCACCTGCTTTATACATTTTTAGGATAAATGTCATACGGTATTTGTCAA 4400
ACCCAAACTTTTTAAATTTTATTTTCAGATCAAAAATTGATGTTAAAAGTTTAAGATATTTACGAAAAAATGTTTTACTTAAAACTTTTTTATATGGATA 4500
AAATTTTAGAACATTCAGATAGGAGTTCCGTCCCTAAAACTTTTGTGGTCGCCGAGAAGCTTTCGAATTAATAATCTCATTTTATAAGTTCGAAACTAAT 4600
TTTTTGTGCAACTGAATTTTTAGAGATGAAACTTTAAGTTTCAATTTATCCCATTTGAAACCGTCCCCTTCTATAAAACTTCAAAATTTTCAGAGTTCAA 4700
CGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAA 4800
TATTATCAACAAGCCTGTTGAGGTGAGTATTTTTTGTTTCTGGGTAGAGGCTTCTTGTCAAGTTTGGCCTAAATTATATTAACTTGGTTTAGAGGCTGGC 4900
AAAGCCATTGATCAAGCATGGGCTAAACTGGGCCGGCCTGAACCATGTACATATTCTTTGGCCCGAGTAGTTGCAATCTAAAGATTCGAAGCTGGCTTCA 5000
AAGTCGGACTAGGCAAAAGTGCAAAAATGGAAAATATCTTGAAATTCAATACGCTTTCCGTCTTTCCATTCTTCCTTTTTTGTCGTCTTTTGTTGCAGAT 5100
TTTCCCTTTTTTAGATTTTAAGATTTTGATACACTTTAATGTCTGGCTTCGCCTTCCTAAGAGCCTTCCTATATTTTCGAAAAATAATCAATTTTTAGGA 5200
AAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATCCACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTT 5300
AGTCCAATTATGGCACATAAGAAGTTGACAAATGGTACGTTTATTTCTGAACTTTACTTATGTTTCGGTCGGTGACGTTTTTGTTGACCATGTGATGGGA 5400
AGTAATTTTTGGATTATTTAAAGTTTGTGCGGGAATAGTAAGGAGGAGTACAATTATTTTATTTGTAGAAGGTTCAGTAAAACTTTGATTTTTCTGACCA 5500
TAAGTTTTTTTTCTGAAAGTTGTTTAAAAAATTGAGTTAAAAAATAAAATAATACTCTATAAAAATTTCTAAATTCTTGGGAATTTTTTTCAAAAATGTT 5600
TTTCCAAATATGTCTAATAGTAAGATTTGTTTGTGAATTTACAAAACATATTTTAAAATACATTTTAATTTATTCGATTTTTTCGTTCGCAGGAACGACA 5700
AAAAATCAGAAAAAGCGAAATTTAATTCCAAAAAAAATATTTTGAAAACTCACAAATAAAGCTACTTTTCAAAAAATCAACAAAAAAAATATCAAAAGAT 5800
TCATATTTTCAGAAATAGAGACATAGTCAAAAACTATCAAAAAATTCACTATTTTCCGGAAAAAAATTGAGAAAAATTCCAAAAATTGTAAAAAACAAAA 5900
TTGAGAAAAATTCCAGAAATTGAAAAAAAATTCCTTTGAGGAAAATTTAAAAATTTTAAATGTGTGATTTCTGAAACCAAGCATTTTCCGACTTTTCGGC 6000

Tuesday, 18 November 1997 11:47 Page 3
fig 33 pEGFPxba (1 > 5447) Site and Sequence

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCA 1800
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT

eGFPC.e.unc53xba

C.e.unc53 xba

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA 1900
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT

eGFPC.e.unc53xba

C.e.unc53 xba

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAA 2000
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT

eGFPC.e.unc53xba

C.e.unc53 xba

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA 2100
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTTGGCTGGATGGAGGGTTGAGGTTTTT

eGFPC.e.unc53xba

C.e.unc53 xba

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q K

CCTTGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAA 2200
GGAACCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTT

eGFPC.e.unc53xba

P W D P P D L D N . S . S A I P H L . R F Y L L . K T S H T S P .

CCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAA 2300
GGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTT

T . N I K . M Q L L L L T C L L Q L I M V T N K A I A S Q I S Q I K

GCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTA 2400
CGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAACATTCGCAATTATAAAACAATTTTAAGCGCAAT

H F F H C I L V V V C P N S S M Y L N A . I V S V N I L L K F A L

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 4

```
AATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTC
TTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAG  2500
 N  F  C  .  I  S  S  F  F  N  Q  .  A  E  I  G  K  I  P  Y  K  S  K  E  .  T  E  I  G  L  S  V  V

CAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACC
GTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGG  2600
P  V  W  N  K  S  P  L  L  K  N  V  D  S  N  V  K  G  R  K  T  V  Y  Q  G  D  G  P  L  R  E  P  S  P

CTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTG
GATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCAC  2700
 .  S  S  F  L  G  S  R  C  R  K  A  L  N  R  N  P  K  G  S  P  R  F  R  A  .  R  G  K  P  A  N  V

GCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATG
CGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGCGGCGCGAATTAC  2800
 A  R  K  E  G  K  K  A  K  G  A  G  A  R  A  L  A  S  V  A  V  T  L  R  V  T  T  T  P  A  A  L  N

CGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
GCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTA  2900
A  P  L  Q  G  A  S  G  G  T  F  R  G  N  V  R  G  T  P  I  C  L  F  F  .  I  H  S  N  M  Y  P  L  M

GAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAA
CTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAGGACTCCGCCTTTCTTGGTCGACACCTTACACACAGTCAATCCCACACCTT  3000
 R  Q  .  P  .  .  M  L  Q  .  Y  .  K  R  K  S  P  E  A  E  R  T  S  C  G  M  C  V  S  .  G  V  E

AGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAG
TCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTC  3100
 S  P  Q  A  P  Q  Q  A  E  V  C  K  A  C  I  S  I  S  Q  Q  P  G  V  E  S  P  Q  A  P  Q  Q  A  E

TATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCC
ATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGG  3200
V  C  K  A  C  I  S  I  S  Q  Q  P  .  S  R  P  .  L  R  P  S  R  P  .  L  R  P  V  P  P  I  L  R  P

CATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCT
GTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGA  3300
 M  A  D  .  F  F  L  F  M  Q  R  P  R  P  P  R  P  L  S  Y  S  R  S  S  E  E  A  F  L  E  A  .  A

TTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTAT
AAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATA  3400
 F  A  K  I  D  Q  E  T  G  .  G  S  F  R  M  I  E  Q  D  G  L  H  A  G  S  P  A  A  W  V  E  R  L

TCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCT
AGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGA  3500
F  G  Y  D  W  A  Q  Q  T  I  G  C  S  D  A  A  V  F  R  L  S  A  Q  G  R  P  V  L  F  V  K  T  D  L

GTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAA
CAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTT  3600
 S  G  A  L  N  E  L  Q  D  E  A  A  R  L  S  W  L  A  T  T  G  V  P  C  A  A  V  L  D  V  V  T  E
```

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 5

```
GCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAA
CGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTT 3700
 A  G  R  D  V  L  L  L  G  E  V  P  G  Q  D  L  L  S  S  H  L  A  P  A  E  K  V  S  I  M  A  D  A

TGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGT
ACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACA 3800
M  R  R  L  H  T  L  D  P  A  T  C  P  F  D  H  Q  A  K  H  R  I  E  R  A  R  T  R  M  E  A  G  L  V

CGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTG
GCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCCGCTCCTAGAGCAGCAC 3900
 D  Q  D  D  L  D  E  E  H  Q  G  L  A  P  A  E  L  F  A  R  L  K  A  S  M  P  D  G  E  D  L  V  V

ACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGG
TGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCC 4000
 T  H  G  D  A  C  L  P  N  I  M  V  E  N  G  R  F  S  G  F  I  D  C  G  R  L  G  V  A  D  R  Y  Q

ACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCG
TGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGC 4100
D  I  A  L  A  T  R  D  I  A  E  E  L  G  G  E  W  A  D  R  F  L  V  L  Y  G  I  A  A  P  D  S  Q  R

CATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCG
GTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGC 4200
 I  A  F  Y  R  L  L  D  E  F  F  .  A  G  L  W  G  S  K  .  P  T  K  R  R  P  T  C  H  H  E  I  S

ATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
TAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAA 4300
 I  P  P  P  P  S  M  K  G  W  A  S  E  S  F  S  G  T  P  A  G  .  S  S  S  A  G  I  S  C  W  S  S

CGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGT
GCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATGGCCTTCCTTGGGCGCGATACTGCCGTTATTTTTCTGTCTTATTTTGCGTGCCA 4400
S  P  T  L  G  G  G  .  L  K  H  G  R  R  Q  Y  R  K  E  P  A  L  .  R  Q  .  K  D  R  I  K  R  T  V

GTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTT
CAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAAGAA 4500
 L  G  R  L  F  I  N  A  G  F  G  P  R  A  G  T  L  S  I  P  H  R  D  P  I  G  A  N  T  P  A  F  L

CCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATAT
GGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTCGGTTGCAGCCCCGCCGTCCGGGACGGTATCGGAGTCCAATGAGTATATA 4600
 P  F  P  H  P  T  P  Q  V  R  V  K  A  Q  G  S  Q  P  T  S  G  R  Q  A  L  P  .  P  Q  V  T  H  I

ACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCG
TGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGC 4700
Y  F  R  L  I  .  N  F  I  F  N  L  K  G  S  R  .  R  S  F  L  I  I  S  .  P  K  S  L  N  V  S  F  R

TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
AAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCG 4800
 S  T  E  R  Q  T  P  .  K  R  S  K  D  L  L  E  I  L  F  F  C  A  .  S  A  A  C  K  Q  K  N  H  R
```

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence Page 6

```
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGT
ATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCA 4900
 Y  Q  R  V  F  V  C  R  I  K  S  Y  Q  L  F  F  R  R  .  L  A  S  A  E  R  R  Y  Q  I  L  S  F  .

GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAG
CATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTC 5000
C  S  R  S  .  A  T  T  S  R  T  L  .  H  R  L  H  T  S  L  C  .  S  C  Y  Q  V  L  L  P  V  A  I  S

TCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAA
AGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTT 5100
 R  V  L  P  G  W  T  Q  D  D  S  Y  R  I  R  R  S  G  R  A  E  R  G  V  R  A  H  S  P  A  W  S  E

CGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGT
GCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCA 5200
 R  P  T  P  N  .  D  T  Y  S  V  S  Y  E  K  A  P  R  F  P  K  G  E  R  R  T  G  I  R  .  A  A  G

CGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTG
GCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAAC 5300
S  E  Q  E  S  A  R  G  S  F  Q  G  E  T  P  G  I  F  I  V  L  S  G  F  A  T  S  D  L  S  V  D  F  C

TGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTC
ACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAG 5400
 D  A  R  Q  G  G  G  A  Y  G  K  T  P  A  T  R  P  F  Y  G  S  W  P  F  A  G  L  L  L  T  C  S  F

CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
GACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 5447
 L  R  Y  P  L  I  L  W  I  T  V  L  P  P  C  I
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence
Enzymes : 100 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Alp

Page 1

```
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC   100
pCMV
 .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  W  P  A  W  L  T

CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA   200
pCMV
 A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S  N  A  N  R  D  F  P  L  T  S  M  G  G  V  F  T  V

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT   300
pCMV
 N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A  P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V

CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT   400
pCMV
 H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H  R  Y  Y  H  G  D  A  V  L  A  V  H  Q  W  A  W

TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT   500
pCMV
 I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q  W  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .

ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT   600
pCMV
 Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G  G  L  Y  K  Q  S  W  F  S  E  P  S  D  P  L  A  L

CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC   700
EGFP
 P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG   800
EGFP
 V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC   900
EGFP
 T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence


```
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 1000
EGFP
R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT 1100
EGFP
D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 1200
EGFP
V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT 1300
EGFP
P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGATAAGCTTGATATCGAATTCCTGCAGCC
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGAGCTCGAGTTCGAAGCTTAAGACGTCAGCTATTCGAACTATAGCTTAAGGACGTCGG 1400
EGFP
I T L G M D E L Y K S G L R S R A Q A S N S A V D K L D I E F L Q P

CCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGC
GGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCG 1500
insert pLM1
ORF pLM1
L L F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S

CTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGT
GACGCTCCCAGGGTCCACTGAGTGTCGAGGGACCTCTACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCA 1600
insert pLM1
ORF pLM1
L R G S Q V T H S S L E M T C Y D S D D A N P R S V S S L S N R S

CCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTG
GGGGAGACAGTACCGCGATACCGGTCAGGTCAGGCGCCGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGAC 1700
insert pLM1
ORF pLM1
S P L S W R Y G Q S S P R L Q A G D A P S V G G S C R S E G T P A W
```

Tuesday, 18 November 1997 11:48 Page 3

fig 34 pLM4 (1 > 10070) Site and Sequence

GTACATGCACGGCGAACGGGCCCACTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTG 1800
CATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGAC

insert pLM1

ORF pLM1

Y M H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L

GACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATG 1900
CTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTAC

insert pLM1

ORF pLM1

D S D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G W D

AAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCC 2000
TTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGG

insert pLM1

ORF pLM1

E S S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P

CACTGCTTCTCGCAGGAACTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAG 2100
GTGACGAAGAGCGTCCTTGAGTTGTTATCACGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTC

insert pLM1

ORF pLM1

T A S R R N S T I V L R T D S E K R S L A E S G L S W F S E S E E

AAAGCCCCTAAAAAACTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCAT 2200
TTTCGGGGATTTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTA

insert pLM1

ORF pLM1

K A P K K L E Y D S G S L K M E P G T S K W R R E R P E S C D D S

CCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCA 2300
GGTTCCCACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGT

insert pLM1

ORF pLM1

S K G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

```
CACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCC
GTGTCGGGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGG  2400
                              insert pLM1
                              ORF pLM1
 T  A  Q  S  A  L  K  V  A  G  K  P  E  G  K  A  T  D  K  G  K  L  A  V  K  N  T  G  L  Q  R  S  S

TCTGATGCTGGTCGGGACCGCCTGAGTGATGCTAAGAAGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTC
AGACTACGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAG  2500
                              insert pLM1
                              ORF pLM1
 S  D  A  G  R  D  R  L  S  D  A  K  K  P  P  S  G  I  A  R  P  S  T  S  G  S  F  G  Y  K  K  P  P

CTGCCACAGGCACAGCCACTGTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCG
GACGGTGTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGC  2600
                              insert pLM1
                              ORF pLM1
 P  A  T  G  T  A  T  V  M  Q  T  G  G  S  A  T  L  S  K  I  Q  K  S  S  G  I  P  V  K  P  V  N  G  R

CAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAG
GTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCGGTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTC  2700
                              insert pLM1
                              ORF pLM1
 K  T  S  L  D  V  S  N  S  A  E  P  G  F  L  A  P  G  A  R  S  N  I  Q  Y  R  S  L  P  R  P  A  K

TCAAGTTCTATGAGCGTGACCGGCGGGCGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGC
AGTTCAAGATACTCGCACTGGCCGCCCGCCCCACCTGGAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCG  2800
                              insert pLM1
                              ORF pLM1
 S  S  S  M  S  V  T  G  G  R  G  G  P  R  P  V  S  S  S  I  D  P  S  L  L  S  T  K  Q  G  G  L  T

CTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGC
GAAGGTCTGACTTCCTCGGATGGTTCCATCGGTCACCCGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCG  2900
                              insert pLM1
                              ORF pLM1
 P  S  R  L  K  E  P  T  K  V  A  S  G  R  T  T  P  A  P  V  N  Q  T  D  R  E  K  E  K  A  K  A  K  A
```

Tuesday, 18 November 1997 11:48 Page 6

fig 34 pLM4 (1 > 10070) Site and Sequence

```
AGTGGCCTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCA 3000
TCACCGGAACCTGAGTCTGTTGTAGAGGAACTTCTCATAACCGAGGGGTCTCTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGT
                                        insert pLM1
                                        ORF pLM1
 V  A  L  D  S  D  N  I  S  L  K  S  I  G  S  P  E  S  T  P  K  N  Q  A  S  H  P  T  A  T  K  L  A

GAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTAC 3100
CTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATG
                                        insert pLM1
                                        ORF pLM1
 E  L  P  P  T  P  L  R  A  T  A  K  S  F  V  K  P  P  S  L  A  N  L  D  K  V  N  S  N  S  L  D  L

CATCATCCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCC 3200
GTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGG
                                        insert pLM1
                                        ORF pLM1
 P  S  S  S  D  T  T  H  A  S  K  V  P  D  L  H  A  T  S  S  A  S  G  G  P  L  P  S  C  F  T  P  S  P

GGCACCCATCCTCAATATTAACTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTC 3300
CCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGAGGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAG
                                        insert pLM1
                                        ORF pLM1
 A  P  I  L  N  I  N  S  A  S  F  S  Q  G  L  E  L  M  S  G  F  S  V  P  K  E  T  R  M  Y  P  K  L

TCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTG 3400
AGTCCGGACGTGTCCTCGTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGAC
                                        insert pLM1
                                        ORF pLM1
 S  G  L  H  R  S  M  E  S  L  Q  M  P  M  S  L  P  S  A  F  P  S  S  T  P  V  P  T  P  P  A  P  P

CTGCTCCCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAA 3500
GACGAGGGTGTCTTCTTCTCTGCCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTT
                                        insert pLM1
                                        ORF pLM1
 A  A  P  T  E  E  E  T  E  E  L  T  V  S  G  S  P  R  A  G  Q  L  D  S  N  Q  R  D  R  N  T  L  P
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

GAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAG
CTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCTCCTCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTC 3600

insert pLM1

ORF pLM1

K G L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E

CTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCT
GACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGA 3700

insert pLM1

ORF pLM1

L P S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R

CCAACAGCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCG
GGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGC 3800

insert pLM1

ORF pLM1

S N S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R

GTCAGGATCCTTCCGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATG
CAGTCCTAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTAC 3900

insert pLM1

ORF pLM1

S G S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M

CAATCTGAGCAAATCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGG
GTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACC 4000

insert pLM1

ORF pLM1

Q S E Q I R K L R R E L E S S Q E K V A T L T S Q L S A N A N L V

CTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAAC
GACGAAAACTCGTCTCGGACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTG 4100

insert pLM1

ORF pLM1

A A F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T

Tuesday, 18 November 1997 13:56 Page 3

fig 53 pLM6 (1 > 4947) Site and Sequence

AGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCA 2400
TCTTCCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGT

U3 stuk

ORF

K K E V S E L R S E L V E K E M K L T D I R L E A L N S A H Q L D Q

GCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACT 2500
CGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGA

U3 stuk

ORF

L R E T M H N M Q L E V D L L K A E N D R L K V A P G P S S G S T

CCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACC 2600
GGTCCCGTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGG

U3 stuk

ORF

P G Q V P G S S A L S S P R R S L G L A L T H S F G P S L A D T D

TGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTT 2700
ACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAA

U3 stuk

ORF

L S P M D G I S T C G P K E E V T L R V V V R M P P Q H I I K G D L

GAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATT 2800
CTTCGTCGTCCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAA

U3 stuk

ORF

K Q Q E F F L G C S K V S G K V D W K M L D E A V F Q V F K D Y I

TCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCG 2900
AGATTTTACCTGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGC

U3 stuk

ORF

S K M D P A S T L G L S T E S I H G Y S I S H V K R V L D A E P P

Tuesday, 18 November 1997 13:57

fig 53 pLM6 (1 > 4947) Site and Sequence

Page 4

AGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCC
TCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGG 3000

U3 stuk

ORF

E M P P C R R G V N N I S V S L K G L K E K C V D S L V F E T L I P

CAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGC
GTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCG 3100

U3 stuk

ORF

K P M M Q H Y I S L L L K H R R L V L S G P S G T G K T Y L T N R

TTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGT
AACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACA 3200

U3 stuk

ORF

L A E Y L V E R S G R E V T E G I V S T F N M H Q Q S C K D L Q L

ATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAG
TAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTC 3300

U3 stuk

ORF

Y L S N L A N Q I D R E T G I G D V P L V I L L D D L S E A G S I S

TGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTG
ACTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAAC 3400

U3 stuk

ORF

E L V N G A L T C K Y H K C P Y I I G T T N Q P V K M T P N H G L

CACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCG
GTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGC 3500

U3 stuk

ORF

H L S F R M L T F S N N V E P A N G F L V R Y L R R K L V E S D S

Tuesday, 18 November 1997 13:57 Page 5
fig 53 pLM6 (1 > 4947) Site and Sequence

ACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGA
TGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCT 3600

U3 stuk

ORF

D I N A N K E E L L R V L D V V P K L V Y H L H T F L E K H S T S D

CTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTAT
GAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATA 3700

U3 stuk

ORF

F L I G P C F F L S C P I G I E D F R T V F I D L W N N S I I P Y

CTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAG
GATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTC 3800

U3 stuk

ORF

L Q E G A K D G I K V H G Q K A A W E D P V E W V R D T L P W P S

CCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCAC
GGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTG 3900

U3 stuk

ORF

A Q Q D Q S K L Y H L P P P T V G P H S I A S P P E D R T V K D S T

CCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGAC
GGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTG 4000

U3 stuk

ORF

P S S L D S D P L M A M L L K L Q E A A N Y I E S P D R E T I L D

CCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCC
GGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGG 4100

U3 stuk

ORF

P N L Q A T L . G F G N H C H P R T A E R V H Q L S . L L L S P L

Tuesday, 18 November 1997 13:57
fig 53 pLM6 (1 > 4947) Site and Sequence

TCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACT 4200
AGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAA

U3 stuk

L F Q S T G S P A P G G E Q E G G G D E R G G T G S W C C T F E N F

CCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAATT 4300
GGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATCTCGCCGGCGGTGGCGCCACCTCGAGGTTAA

U3 stuk

L G R N G G V A F G N L C P L N T F T G L L . S G R H R G G A P I

CGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGC 4400
GCGGGATATCACTCAGCATAATGCGCGCGAGTGACCGGCAGCAAAATGTTGCAGCACTGACCCTTTTGGGACCGCAATGGGTTGAATTAGCGGAACGTCG

R P I V S R I T R A H W P S F Y N V V T G K T L A L P N L I A L Q

ACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGT 4500
TGTAGGGGGAAAGCGGTCGACCGCATTATCGCTTCTCCGGGCGTGGCTAGCGGGAAGGGTTGTCAACGCGTCGGACTTACCGCTTACCCTGCGCGGGACA

H I P L S P A G V I A K R P A P I A L P N S C A A . M A N G T R P V

AGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCT 4600
TCGCCGCGTAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGA

A A H . A R R V W W L R A A . P L H L P A P . R P L L S L S S L P

TTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACT 4700
AAGAGCGGTGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCCGAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGA

F S P R S P A F P V K L . I G G S L . G S D L V L Y G T S T P K N

TGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTC 4800
ACTAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAG

L I R V M V H V V G H R P D R R F F A L . R V S P R S L I V D S C S

CAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAAC 4900
GTTTGACCTTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACGGCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTG

K L E Q H S T L S R S I L L I Y K G F C R F R P I G . K M S . F N

AAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAG 4947
TTTTTAAATTGCGCTTAAAATTGTTTTATAATTGCGAATGTTAAATC

K N L T R I L T K Y . R L Q F R

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence
Enzymes : 72 of 148 enzymes (Filtered)
Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

```
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
CACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTA 100
 G  G  T  F  R  G  N  V  R  G  T  P  I  C  L  F  F  .  I  H  S  N  M  Y  P  L  M  R  Q  .  P  .  .  M

GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTG 200
  L  Q  .  Y  .  K  R  K  S  M  S  I  Q  H  F  R  V  A  L  I  P  F  F  A  A  F  C  L  P  V  F  A  H

CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
GGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAA 300
 P  E  T  L  V  K  V  K  D  A  E  D  Q  L  G  A  R  V  G  Y  I  E  L  D  L  N  S  G  K  I  L  E  S

TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
AAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTCGTTGAGCCAGC 400
F  R  P  E  E  R  F  P  M  M  S  T  F  K  V  L  L  C  G  A  V  L  S  R  I  D  A  G  Q  E  Q  L  G  R

CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
GGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGG 500
 R  I  H  Y  S  Q  N  D  L  V  E  Y  S  P  V  T  E  K  H  L  T  D  G  M  T  V  R  E  L  C  S  A  A

ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA
TATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATT 600
 I  T  M  S  D  N  T  A  A  N  L  L  L  T  T  I  G  G  P  K  E  L  T  A  F  L  H  N  M  G  D  H  V

CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT
GAGCGGAACTAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGA 700
T  R  L  D  R  V  E  P  E  L  N  E  A  I  P  N  D  E  R  D  T  T  M  P  V  A  M  A  T  T  L  R  K  L

ATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCG
TAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGC 800
  L  T  G  E  L  L  T  L  A  S  R  Q  Q  L  I  D  V  M  E  A  D  K  V  A  G  P  L  L  R  S  A  L  P

GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG
CGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATC 900
 A  G  V  F  I  A  D  K  S  G  A  G  E  R  G  S  R  G  I  I  A  A  L  G  P  D  G  K  P  S  R  I  V

TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCA
AATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGT 1000
V  I  Y  T  T  G  S  Q  A  T  M  D  E  R  N  R  Q  I  A  E  I  G  A  S  L  I  K  H  V  .  L  S  D  Q

AGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCT
TCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGA 1100
  V  Y  S  Y  I  L  .  I  D  L  K  L  H  F  .  F  K  R  I  .  V  K  I  L  F  D  N  L  M  T  K  I  P

TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA
ATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTT 1200
 .  R  E  F  S  F  H  .  A  S  D  P  V  E  K  I  K  G  S  S  .  D  P  F  F  L  R  V  I  C  C  L  Q

CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAA
GTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTT 1300
T  K  K  P  P  L  P  A  V  V  C  L  P  D  Q  E  L  P  T  L  F  P  K  V  T  G  F  S  R  A  Q  I  P  N

TACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
ATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGA 1400
  T  V  L  L  V  .  P  .  L  G  H  H  F  K  N  S  V  A  P  P  T  Y  L  A  L  L  I  L  L  P  V  A  A
```

Tuesday, 18 November 1997 13:57

fig 54 pLM1 (1 > 8285) Site and Sequence

Page 2

GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC

CGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCG 1500

A S G D K S C L T G L D S R R . L P D K A Q R S G . T G G S C T Q

CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC

GGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGG 1600

P S L E R T T Y T E L R Y L Q R E L . E S A T L P E G R K A D R Y P

GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTT

CCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAA 1700

V S G R V G T G E R T R E L P G G N A V Y L Y S P V G F R H L . L

GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG

CTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAAC 1800

E R R F L . C S S G G R S L V K N A S N A A F L R F L A F C V P F

CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCG

GAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGC 1900

A H M F F P A L S P D S V D N R I T A F E . A D T A R R S R T T E R

CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTT

GTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAA 2000

S E S V S E E A E E R P I R K P P L P A R V P I H . C S V H D R F

CCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT

GGGCTGACCTTTCGCCCGTCACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACA 2100

P D V K A G S E R N A I N V S . L T H . A P Q A L H F M L P A R M

TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCT

ACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATACTGGTACTAATGCGGTTCGCGCGTTAATTGGGAGTGATTTCCCTTGTTTTCGA 2200

L C G I V S G . Q F H T G N S Y D H D Y A K R A I N P H . R E Q K L

GGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAG

CCCATGGCCCGGGGGGGAGCTCCAGCTGCCATAGCTATTCGAACTATAGCTTAAGGACGTCGGGGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTC 2300

insert pLM1

ORF pLM1

G T G P P L E V D G I D K L D I E F L Q P L L F S Q M L D P E S Q

AGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGA

TCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCGGACGCTCCCAGGGTCCACTGAGTGTCGAGGGACCTCT 2400

insert pLM1

ORF pLM1

R K R T V Q N V L D L R Q N L E E T M S S L R G S Q V T H S S L E

TGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCGCG

ACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAGACAGTACCGCGATACCGGTCAGGTCAGGCGC 2500

insert pLM1

ORF pLM1

M T C Y D S D D A N P R S V S S L S N R S S P L S V R Y G Q S S P R

Tuesday, 18 November 1997 13:57 Page 3

fig 54 pLM1 (1 > 8285) Site and Sequence

GCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTACATGCACGGCGAACGGGCCCACTACTCCCACACC 2600
CGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGG

insert pLM1

ORF pLM1

L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H T

ATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGA 2700
TACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTACT

insert pLM1

ORF pLM1

M P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y M

GCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCTC 2800
CGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTACTTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAG

insert pLM1

ORF pLM1

S D S D L M G K T M T E D D D I T T G W D E S S S I S S G L S D A S

AGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGCAGGAACTCAACAATAGTGCTACGC 2900
TCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACGAAGAGCGTCCTTGAGTTGTTATCACGATGCG

insert pLM1

ORF pLM1

D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L R

ACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCCTAAAAAACTGGAGTACGACAGTGGTAGCC 3000
TGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTCTTTCGGGGATTTTTTGACCTCATGCTGTCACCATCGG

insert pLM1

ORF pLM1

T D S E K R S L A E S G L S W F S E S E E K A P K K L E Y D S G S

TGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCATCCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGG 3100
ACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCCCACCTCTTGACTTTTTCGGGTAGTCGGACCC

insert pLM1

ORF pLM1

L K M E P G T S K W R R E R P E S C D D S S K G G E L K K P I S L G

CCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCACACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAACCT 3200
GGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCGGGTCTCACGGGAGTTTCAGCGTCCGTTTGGA

insert pLM1

ORF pLM1

H P G S L K K G K T P P V A V T S P I T H T A Q S A L K V A G K P

GAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGATGCTGGTCGGGACCGCCTGAGTGATGCTAAGA 3300
CTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGGAGACTACGACCAGCCCTGGCGGACTCACTACGATTCT

insert pLM1

ORF pLM1

E G K A T D K G K L A V K N T G L Q R S S S D A G R D R L S D A K

Tuesday, 18 November 1997 13:57 Page 4

fig 54 pLM1 (1 > 8285) Site and Sequence

```
AGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTCCTGCCACAGGCACAGCCACTGTCATGCAAACTGGTGG
TCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAGGACGGTGTCCGTGTCGGTGACAGTACGTTTGACCACC 3400
insert pLM1
ORF pLM1
K P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G G

TTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCA
AAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCGGT 3500
insert pLM1
ORF pLM1
S A T L S K I Q K S S G I P V K P V N G R K T S L D V S N S A E P

GGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGTTCTATGAGCGTGACCGGCGGGCGGGGTGGAC
CCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCAAGATACTCGCACTGGCCGCCCGCCCCACCTG 3600
insert pLM1
ORF pLM1
G F L A P G A R S N I Q Y R S L P R P A K S S S M S V T G G R G G

CTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGG
GAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGGTCTGACTTCCTCGGATGGTTCCATCGGTCACC 3700
insert pLM1
ORF pLM1
P R P V S S S I D P S L L S T K Q G G L T P S R L K E P T K V A S G

GCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGAGT
CGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCGTCACCGGAACCTGAGTCTGTTGTAGAGGAACTTCTCA 3800
insert pLM1
ORF pLM1
R T T P A P V N Q T D R E K E K A K A K A V A L D S D N I S L K S

ATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGA
TAACCGAGGGGTCTCTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGTCTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTTCT 3900
insert pLM1
ORF pLM1
I G S P E S T P K N Q A S H P T A T K L A E L P P T P L R A T A K

GCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAAGGTCCC
CGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATGGTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGG 4000
insert pLM1
ORF pLM1
S F V K P P S L A N L D K V N S N S L D L P S S S D T T H A S K V P

AGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGCACCCATCCTCAATATTAACTCAGCCAGCTTCTCC
TCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGAGG 4100
insert pLM1
ORF pLM1
D L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F S
```

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 5

CAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGC
GTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCGGACGTGTCCTCGTACCTCAGGGAGGTCTACG 4200

insert pLM1

ORF pLM1

Q G L E L M S G F S V P K E T R M Y P K L S G L H R S M E S L Q M

CAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTCCCACAGAAGAAGAGACGGAAGAGCTGACTTG
GTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGTCTTCTTCTCTGCCTTCTCGACTGAAC 4300

insert pLM1

ORF pLM1

P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T V

GAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAG
CTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCTCCTC 4400

insert pLM1

ORF pLM1

S G S P R A G Q L D S N Q R D R N T L P K K G L R Y Q L Q S Q E E

ACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAGTG
TGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCAC 4500

insert pLM1

ORF pLM1

T K E R R H S H T I G G L P E S D D Q S E L P S P P A L P M S L S

CAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACAGCATCCCCACCCACGAGGCGGCCTTCGAGCT
GTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTCGA 4600

insert pLM1

ORF pLM1

A K G Q L T N I V S P T A A T T P R I T R S N S I P T H E A A F E L

GTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGGATCCTTCCGAGACCCCACGGACGATGTTCAC
CATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCCTAGGAAGGCTCTGGGGTGCCTGCTACAAGTG 4700

insert pLM1

ORF pLM1

Y S G S Q M G S T L S L A E R P K G M I R S G S F R D P T D D V H

GGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAAATCCGGAAGCTTCGTAGGGAACTGG
CCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACC 4800

insert pLM1

ORF pLM1

G S V L S L A S S A S S T Y S S A E E R M Q S E Q I R K L R R E L

AATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCG
TTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAAAACTCGTCTCGGACCACTTATACTGTAGGGC 4900

insert pLM1

ORF pLM1

E S S Q E K V A T L T S Q L S A N A N L V A A F E Q S L V N M T S R

CCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAG
GGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTTCTTGAGACTCCGGGTC 5000

insert pLM1

ORF pLM1

L R H L A E T A E E K D T E L L D L R E T I D F L K K K N S E A Q

GCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCA
CGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCTGTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGT 5100

insert pLM1

ORF pLM1

A V I Q G A L N A S E T T P K E L R I K R Q N S S D S I S S L N S

TCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGC
AGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCG 5200

insert pLM1

ORF pLM1

I T S H S S I G S S K D A D A K K K K K K S W V Y E L R S S F N K A

GTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCCAAACTACAG
CAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTC 5300

insert pLM1

ORF pLM1

F S I K K G P K S A S S Y S D I E E I A T P D S S A P S S P K L Q

CATGGTTCCACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCACA
GTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGTGT 5400

insert pLM1

ORF pLM1

H G S T E T A S P S I K S S T L S S V G T D V T E G P A H P A P H

CTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACAT
GATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCCTTTACTTCGAATGTCTGTA 5500

insert pLM1

ORF pLM1

T R L F H A N E E E E P E K K E V S E L R S E L V E K E M K L T D I

CCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGAC
GGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTG 5600

insert pLM1

ORF pLM1

R L E A L N S A H Q L D Q L R E T M H N M Q L E V D L L K A E N D

CGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCAC
GCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTG 5700

insert pLM1

ORF pLM1

R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S L G L A

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

```
TCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGT
AGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACTGGGAGGCCCACCA 5800
insert pLM1
ORF pLM1
L T H S F G P S L A D T D L S P M D G I S T C G P K E E V T L R V V

GGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATG
CCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCTAC 5900
insert pLM1
ORF pLM1
V R M P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D W K M

CTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCA
GACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGT 6000
insert pLM1
ORF pLM1
L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I H G Y S

TCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGA
AGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCT 6100
insert pLM1
ORF pLM1
I S H V K R V L D A E P P E M P P C R R G V N N I S V S L K G L K E

GAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCG
CTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGC 6200
insert pLM1
ORF pLM1
K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S

GGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCT
CCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGA 6300
insert pLM1
ORF pLM1
G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T

TCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGT
AGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCA 6400
insert pLM1
ORF pLM1
F N M H Q Q S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L V

GATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACC
CTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCATGG 6500
insert pLM1
ORF pLM1
I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y I I G T
```

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 9

ACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGG
TGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACC 6600

insert pLM1

ORF pLM1

T N Q P V K M T P N H G L H L S F R M L T F S N N V E P A N G F L

TTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTA
AAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACCAT 6700

insert pLM1

ORF pLM1

V R Y L R R K L V E S D S D I N A N K E E L L R V L D W V P K L V Y

TCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACC
AGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGG 6800

insert pLM1

ORF pLM1

H L H T F L E K H S T S D F L I G P C F F L S C P I G I E D F R T

TGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACC
ACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGG 6900

insert pLM1

ORF pLM1

W F I D L W N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D

CAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCAT
GTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTA 7000

insert pLM1

ORF pLM1

P V E W V R D T L P W P S A Q Q D Q S K L Y H L P P P T V G P H S I

TGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCC
ACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGG 7100

insert pLM1

ORF pLM1

A S P P E D R T V K D S T P S S L D S D P L M A M L L K L Q E A A

AACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAAC
TTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTG 7200

insert pLM1

ORF pLM1

N Y I E S P D R E T I L D P N L Q A T L . G F G N H C H P R T A E

GCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAG
CGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTTTC 7300

insert pLM1

R W H Q L S . L L L S P L L F Q S T G S P A P G G E Q E G G G D E R

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

AGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGC
TCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGACCG 7400

insert pLM1

G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T G

CTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAA
GAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTTTT 7500

insert pLM1

L L . . L V G K D D S G S F P . L L V S I T N S V A F V G G V Q K

CATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGGCTGAACTTGCTCAAAAGAAGCCGAATTCCAGCACACTGGCGGCCGTTACT
GTAGTTTTGTGACGTCGTCAAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGTTTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAATGA 7600

insert pLM1

T S K H C S S S P E F S L D L T R L N L L K R S R I P A H V R P L L

AGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGG
TCAAGATCTCGCCGGCGGTGGCGCCACCTCGAGGTTAAGCGGGATATCACTCAGCATAATGCGCGCGAGTGACCGGCAGCAAAATGTTGCAGCACTGACC 7700

V L E R P P P R V S S N S P Y S E S Y Y A R S L A V V L Q R R D V

GAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAAC
CTTTTGGGACCGCAATGGGTTGAATTAGCGGAACGTCGTGTAGGGGGAAAGCGGTCGACCGCATTATCGCTTCTCCGGGCGTGGCTAGCGGGAAGGGTTG 7800
E N P G V T Q L N R L A A H P P F A S V R N S E E A R T D R P S Q

AGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
TCAACGCGTCGGACTTACCGCTTACCCTGCGCGGGACATCGCCGCGTAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTC 7900
Q L R S L N G E V D A P C S G A L S A A G V V V T R S V T A T L A S

CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTC
GCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCCGAGGGAAATCCCAAG 8000
A L A P A P F A F F P S F L A T F A G F P R Q A L N R G L P L G F

CGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGA
GCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACT 8100
R F S A L R H L D P K K L D . G D G S R S G P S P . . T V F R P L

CGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCC
GCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACCTTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACGG 8200
T L E S T F F N S G L L F Q T G T T L N P I S V Y S F D L . G I L P

GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAG
CTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATTGCGCTTAAAATTGTTTTATAATTGCGAATGTTAAATC 8285
I S A Y V L K N E L I . Q K F N A N F N K I L T L T I .

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence
Enzymes : All 146 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

```
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA  100
 T  D  R  E  I  S  R  S  P  M  V  D  S  Q  Y  N  L  L  .  C  R  I  V  K  P  V  S  A  P  C  L  C  V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACTGGCTGTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGC  200
 G  G  R  .  V  V  R  E  Q  N  L  S  Y  N  K  A  R  L  D  R  Q  L  H  E  E  S  A  .  G  .  A  F  C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGGTCTATATGCGCAACTGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT  300
 A  A  S  R  C  T  G  Q  I  Y  A  L  T  L  I  I  D  .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCA  400
 G  V  P  R  Y  I  T  Y  G  K  W  P  A  W  L  T  A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG  500
 N  A  N  R  D  F  P  L  T  S  M  G  G  L  F  T  V  N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A

CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT  600
 P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V  H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTT  700
 R  Y  Y  H  G  D  A  V  L  A  V  H  Q  W  A  V  I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q

TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG
ACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTC  800
 W  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .  Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG  900
                                                                   T7 promotor priming site
 G  L  Y  K  Q  S  S  L  A  N  .  R  T  H  C  L  L  A  Y  R  N  .  Y  D  S  L  .  G  D  P  S  V  L  A

GTTTAAACTTAAGCTTACCATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG  1000
                                      Probond binding domain
 F  K  L  K  L  T  M  G  G  S  H  H  H  H  H  H  G  M  A  S  M  T  G  G  Q  Q  M  G  R  D  L  Y  D

GATGACGATAAGGTACCCGGATCCTTCCGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAG
CTACTGCTATTCCATGGGCCTAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTC  1100
                                      pCB251 insert = U2
 D  D  D  K  V  P  G  S  F  R  D  P  T  D  D  V  H  G  S  V  L  S  L  A  S  S  A  S  S  T  Y  S  S
```

Tuesday, 18 November 1997 13:57 Page 2

fig 55 pCB251 (1 > 8197) Site and Sequence

CTGAGGAGAGGATGCAATCTGAGCAAATCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAA 1200
GACTCCTCTCCTACGTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTT

pCB251 insert = U2

U2 ORF

A E E R M Q S E Q I R K L R R E L E S S Q E K V A T L T S Q L S A N

TGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTG 1300
ACGATTAGACCACCGACGAAAACTCGTCTCGGACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGAC

pCB251 insert = U2

U2 ORF

A N L V A A F E Q S L V N M T S R L R H L A E T A E E K D T E L L

GATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAAC 1400
CTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTG

pCB251 insert = U2

U2 ORF

D L R E T I D F L K K K N S E A Q A V I Q G A L N A S E T T P K E

TTCGGATCAAGAGACAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAA 1500
AAGCCTAGTTCTCTGTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTT

pCB251 insert = U2

U2 ORF

L R I K R Q N S S D S I S S L N S I T S H S S I G S S K D A D A K K

GAAGAAAAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATA 1600
CTTCTTTTTTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATAT

pCB251 insert = U2

U2 ORF

K K K K S W V Y E L R S S F N K A F S I K K G P K S A S S Y S D I

GAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGT 1700
CTCCTCTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGATGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACA

pCB251 insert = U2

U2 ORF

E E I A T P D S S A P S S P K L Q H G S T E T A S P S I K S S T L

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 3

CCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGT
GGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCA 1800

pCB251 insert = U2

U2 ORF

S S V G T D V T E G P A H P A P H T R L F H A N E E E E P E K K E V

ATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAG
TAGCCTCGACGCGAGACTCGATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTC 1900

pCB251 insert = U2

U2 ORF

S E L R S E L W E K E M K L T D I R L E A L N S A H Q L D Q L R E

ACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGG
TGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCC 2000

pCB251 insert = U2

U2 ORF

T M H N M Q L E V D L L K A E N D R L K V A P G P S S G S T P G Q

TCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCAT
AGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTA 2100

pCB251 insert = U2

U2 ORF

V P G S S A L S S P R R S L G L A L T H S F G P S L A D T D L S P M

GGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAG
CCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTC 2200

pCB251 insert = U2

U2 ORF

D G I S T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q

GAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGG
CTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACC 2300

pCB251 insert = U2

U2 ORF

E F F L G C S K V S G K V D W K M L D E A V F Q V F K D Y I S K M

Tuesday, 18 November 1997 13:57 Page 4
fig 55 pCB251 (1 > 8197) Site and Sequence

```
ACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCC
TGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGG  2400
                              pCB251 insert = U2
                                   U2 ORF
D  P  A  S  T  L  G  L  S  T  E  S  I  H  G  Y  S  I  S  H  V  K  R  V  L  D  A  E  P  P  E  M  P  F

TTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATG
AACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTAC  2500
                              pCB251 insert = U2
                                   U2 ORF
 C  R  R  G  V  N  N  I  S  V  S  L  K  G  L  K  E  K  C  V  D  S  L  V  F  E  T  L  I  P  K  P  M

ATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGT
TACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCA  2600
                              pCB251 insert = U2
                                   U2 ORF
M  Q  H  Y  I  S  L  L  L  K  H  R  R  L  V  L  S  G  P  S  G  T  G  K  T  Y  L  T  N  R  L  A  E

ACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAA
TGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTT  2700
                              pCB251 insert = U2
                                   U2 ORF
Y  L  V  E  R  S  G  R  E  V  T  E  G  I  V  S  T  F  N  M  H  Q  Q  S  C  K  D  L  Q  L  Y  L  S  N

CCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTC
GGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAG  2800
                              pCB251 insert = U2
                                   U2 ORF
 L  A  N  Q  I  D  R  E  T  G  I  G  D  V  P  L  V  I  L  L  D  D  L  S  E  A  G  S  I  S  E  L  V

AATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCT
TTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGA  2900
                              pCB251 insert = U2
                                   U2 ORF
N  G  A  L  T  C  K  Y  H  K  C  P  Y  I  I  G  T  T  N  Q  P  V  K  M  T  P  N  H  G  L  H  L  S
```

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

TCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGC
AGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACG 3000

pCB251 insert = U2

U2 ORF

F R M L T F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A

CAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATC
GTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAG 3100

pCB251 insert = U2

U2 ORF

N K E E L L R V L D W V P K L W Y H L H T F L E K H S T S D F L I

GGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAG
CCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTC 3200

pCB251 insert = U2

U2 ORF

G P C F F L S C P I G I E D F R T W F I D L W N N S I I P Y L Q E

GAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGA
CTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCT 3300

pCB251 insert = U2

U2 ORF

G A K D G I K V H G Q K A A W E D P V E W V R D T L P W P S A Q Q D

CCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCT
GGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGA 3400

pCB251 insert = U2

U2 ORF

Q S K L Y H L P P P T V G P H S I A S P P E D R T V K D S T P S S

CTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTC
GACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAG 3500

pCB251 insert = U2

U2 ORF

L D S D P L M A M L L K L Q E A A N Y I E S P D R E T I L D P N L

AGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAG 3600
TCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTC

pCB251 insert = U2

U2 ORF

Q A T L . G F G N H C H P R T A E R W H Q L S . L L L S P L L F Q S

CACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGG 3700
GTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCC

pCB251 insert = U2

T G S P A P G G E Q E G G G D E R G G T G S V C C T F E N F L G R

AATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCT 3800
TTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATCTCCCGGGCAAATTTGGGCGACTAGTCGGAGCTGACACGGA

pCB251 insert = U2

N G G V A F G N L C P L N T F T G L L . R A R L N P L I S L D C A

TCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTG 3900
AGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCACGGAAGGAACTGGGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTTACTCCTTTAAC

F . L P A I C C L P L P R A F L D P G R C H S H C P F L I K . G N C

CATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA 4000
GTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCCACCCCACCCCGTCCTGTCGTTCCCCCTCCTAACCCTTCTGTTATCGTCCGTACGACCCCT

I A L S E . V S F Y S G G W G G A G Q Q G G G L G R Q . Q A C W G

TGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTG 4100
ACGCCACCCGAGATACCGAAGACTCCGCCTTTCTTGGTCGACCCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCGTAATTCGCGCCGCCCACAC

C G G L Y G F . G G K N Q L G L . G V S P R A L . R R I K R G G C

GTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCC 4200
CACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGG

G G Y A Q R D R Y T C Q R P S A R S F R F L P F L S R H V R R L S P

GTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGG 4300
CAGTTCGAGATTTAGCCCCGTAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTGCATCACC

S S S K S G H P F R V P I . C F T A P R P Q K T . L G . W F T . W

GCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATC 4400
CGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACCTTGTTGTGAGTTGGGATAG

A I A L I D G F S P F D V G V H V L . . W T L V P N W N N T Q P Y

TCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTG 4500
AGCCAGATAAGAAAACTAAATATTCCCTAAAACCCCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATTGCGCTTAATTAAGACAC

L G L F F . F I R D F G D F G L L V K K . A D L T K I . R E L I L W

Tuesday, 18 November 1997 13:57 Page 7
fig 55 pCB251 (1 > 8197) Site and Sequence

GAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAA 4600
CTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTT
N V C Q L G C G K S P G S P G R Q K Y A K H A S Q L V S N Q V W K

GTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT 4700
CAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGA
V P R L P S R Q K Y A K H A S Q L V S N H S P A P N S A H P A P N

CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGT 4800
GGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCA
S A Q F R P F S A P W L T N F F Y L C R G R G R L C L . A I P E V V

GAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCG 4900
CTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTCGAGGGCCCTCGAACATATAGGTAAAAGCCTAGACTAGTTCTCTGTCCTACTCCTAGCAAAGC
R R L F W R P R L L Q K A P G S L Y I H F R I . S R D R M R I V S

CATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGAT 5000
GTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTA
H D . T R W I A R R F S G R L G G E A I R L . L G T T D N R L L .

GCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGC 5100
CGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCCTGCTCCGTCGCGCCG
C R R V P A V S A G A P G S F C Q D R P V R C P E . T A G R G S A A

TATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCA 5200
ATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGT
I V A G H D G R S L R S C A R R C H . S G K G L A A I G R S A G A

GGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTC 5300
CCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAG
G S P V I S P C S C R E S I H H G . C N A A A A Y A . S G Y L P I

GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC 5400
CTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCG
R P P S E T S H R A S T Y S D G S R S C R S G . S G R R A S G A R A

CAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA 5500
GTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTT
S R T V R Q A Q G A H A R R R G S R R D P W R C L L A E Y H G G I

TGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGC 5600
ACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCG
W P L F W I H R L W P A G C G G P L S G H S V G Y P . Y C . R A W

GGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGG 5700
CCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCC
R R M G . P L P R A L R Y R R S R F A A H R L L S P S . R V L L S G

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 8

GACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATC
CTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAG 5800
T L G F E M T D Q A T P N L P S R D F D S T A A F Y E R L G F G I

GTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACA
CAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTCGAATATTACCAATGT 5900
V F R D A G W M I L Q R G D L M L E F F A H P N L F I A A Y N G Y

AATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTG
TTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGAC 6000
K . S N S I T N F T N K A F F S L H S S C G L S K L I N V S Y H V C

TATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCC
ATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTTGTATGCTCGG 6100
I P S T S S . S L A . S W S . L F P V . N C Y P L T I P H N I R A

GGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
CCTTCGTATTTCACATTTCGGACCCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCA 6200
G S I K C K A V G A . . V S . L T L I A L R S L P A F Q S G N L S

GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
CGGTCGACGTAATTACTTAGCCGGTTGCGCGCCCCTCTCCGCCAAACGCATAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCA 6300
C Q L H . . I G Q R A G R G G L R I G R S S A S S L T D S L R S V V

TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAG
AGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTCCGGTC 6400
R L R R A V S A H S K A V I R L S T E S G D N A G K N M . A K G Q

CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCA 6500
Q K A R N R K K A A L L A F F H R L R P P D E H H K N R R S S Q R

GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
CCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAG 6600
W R N P T G L . R Y Q A F P P G S S L V R S P V P T L P L T G Y L S

CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCAC
GCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTG 6700
A F L P S G S V A L S Q C S R C R Y L S S V . V V R S K L G C V H

GAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
CTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGAC 6800
E P P V Q P D R C A L S G N Y R L E S N P V R H D L S P L A A T

GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
CATTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGAC 6900
G N R I S R A R Y V G G A T E F L K V V P N Y G Y T R R T V F G I C

Tuesday, 18 November 1997 13:57 Page 1
fig 55 pCB251 (1 > 8197) Site and Sequence

CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG 7000
GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTC
A L L K P V T F G K R V G S S . S G K Q T T A G S G G F F V C K Q

CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG 7100
GTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACC
Q I T R R K K G S Q E D P L I F S T G S D A Q W N E N S R . G I L

TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAG 7200
AGTACTCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTC
V M R L S K R I F T . I L L N . K . S F K S I . S I Y E . T W S D S

TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGG 7300
AATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCC
Y Q C L I S E A P I S A I C L F R S S I V A . L P V V . I T T I R

GAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCG 7400
CTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGC
E G L P S G P S A A M I P R D P R S P A P D L S A I N Q P A G R A

AGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA 7500
TCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTT
E R R S G P A T L S A S I Q S I N C C R E A R V S S S P V N S L R N

CGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC 7600
GCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGG
V V A I A T G I V V S R S S F G M A S F S S G S Q R S R R V T . S

CCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC 7700
GGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACG
P M L C K K A V S S F G P P I V V R S K L A A V L S L M V M A A L

ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG 7800
TATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAAC
H N S L T V M P S V R C F S V T G E Y S T K S F . E . C M R R P S C

CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG 7900
GAGAACGGGCCGCAGTTATGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCC
S C P A S I R D N T A P H S R T L K V L I I G K R S S G R K L S R

ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAA 8000
TAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTT
I L P L L R S S S M . P T R A P N . S S A S F T F T S V S G . A K

CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG 8100
GTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCC
T G R Q N A A K K G I R A T R K C . I L I L F L F Q Y Y . S I Y Q G

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 10

```
TTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC  8197
AATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAG
 Y  C  L  M  S  G  Y  I  F  E  C  I  .  K  N  K  Q  I  G  V  P  R  T  F  P  R  K  V  P  P  D  V
```

Figure 56

Monday, 1 December 1997 14:13 Page 1
fig56 pNP10

```
       10        20        30        40        50        60        70
AAGCTTGCATGCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAGAAAT  70
TGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAATTTTGAGTTTACGACTA  140
CAAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCTATTCTGTATGACTAGTTGTTGAGTGATTTT  210
TCATTGAGAAAATATTAAAAGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTTAGTTTTTCC  280
ATCAACTAGATCTTACAAAACTTGCAATACAATTCCATTTTCAGATTACCCTCGCCACGTGTCGCCACGT  350
      360       370       380       390       400       410       420
CAGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCC  420
ACAGTCAAGAATATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGTATAAAATGCCCAGAAAA  490
CAAGATGATAAAAATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTCTAATATTGACGGCTTATATCT  560
CAATTTTCTTGAGTTTTATCAAAAAATTTTCCACTATACAAATGTAGAAAAGTATTTTTGCACAAATTTTG  630
TCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAAAGTTGGAAGGAGCGCAAG  700
      710       720       730       740       750       760       770
TCTATACTGGAAATAATGATCTGAAACAAATTTGTGCTATTCTCAAATGTTTAAGACATGTTTTGAAGAT  770
TTTTTCAAATTCGCACTAGTTTCAGAACCTTCCTTTTTGTATGAAAAAGTAAAAAAAAACTATTTCAAAC  840
CCTCACCGCCACCATGTTTCAACTCTTAATTTTTATAAAATTTTGCAATTTACAAATCGCCTCCCCTTGC  910
CCGAAAAGTGCCCACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATTGATTGTGAGAAAACACAG  980
AAGAGGCTAACTAAATTGACAGGGACAGGTTGTCCCTCTTCTCCCTCCTTCTCCCGCCTCCTCCTCCGT  1050
      1060      1070      1080      1090      1100      1110      1120
TCCATCTCCAACAACAACAATTTTTCCAATTTCGTTGTCCATTTTGCTTATAAACATTTGTGTGTCGAAGG  1120
AAACTACACGGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACTCTTCGGAAATTGATGAA  1190
TAAAGATAGAGCCGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTTTTTCATCGTCTCAACTT  1260
GCGCATGAGTCTTCCCCCGCTCTCATCACTGACAATTAATGTCGGGTTTTATGCGCTCTTTCCTATTCCG  1330
CCACTCATTCTGGCTTACCACAAACTGGAATACATTTTACTACTATTCAAGCCATTTATTTTGATATTTA  1400
      1410      1420      1430      1440      1450      1460      1470
ATTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACGATATTTTCGATTTTAAA  1470
AAATCTGAAAAGTTTCAAAAAATCAATAAATATTCCCTAACAAATTGTATGGCTAAAATTTTATTTCTAC  1540
TGTTGACAATATCTTTATATGTATCACTGTTTTCCATCTCAAAACCTTGAATCCCCCAAGTTATAGGAAG  1610
CTCCGTGTCACATTTCCCATGCTATGAATCGCTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGA  1680
TAATTATTGGGCACGCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAAGCACAGATTATCTATCAAA  1750
      1760      1770      1780      1790      1800      1810      1820
TTCAATCTTTGAACATTCAGCCAGTTCGTACAATTTTCCATGCTTTTTGGCCCATTAAAAAACTTTCTCA  1820
CCTCTTCATCCATCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACTCTACTTTTTAAGAGAAGGA  1890
GATACTGAGCCACATGGCGTGTGACCCTTTTCATCTCGTCCGTTCGGTCTCAAATTCACGCTCATACTAA  1960
CTCTTCAAATAGCCATAGACCTCCTTGTTTTCTTCTTCGTTTTGACTCGCGCCTATTTTTTGTGGCTGCC  2030
TGAAAGCCGGGAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATAAAAAACGAGGCAATTTA  2100
      2110      2120      2130      2140      2150      2160      2170
AAAATATTAAAATTAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAAACAAAACAAATAGGAAC  2170
CGCCAGATCAAAATTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGGCTCAACAGAAAACTGAA  2240
GTGCCTGCATAAATCTAGTGTAACGTTTAGATTGAACTCGGAAATCCTAAGCCTGAACTATAGCCTTATT  2310
CTAGATCTTAGTTGCGCATAAGCTCAAGCCCAACCAGAAATGACTTGCATTTAGTTTAAGCCTAGATTGA  2380
CTTGCTTGCTTCAGTCTAATCCAGACTAGATTTCCAAGAGAGTTTTCAATTTTAAATGTTTCCAGTTTCT  2450
```

Figure 56

Monday, 1 December 1997 14:13 Page 2
fig 56 pNP10

```
      2460      2470      2480      2490      2600      2510      2520
TGTTACTTAAAATCTTAATGCCCTGTGATGCGTAAAATCGTTATCCCTTTCTCTCACACTTTCAATTACA 2520
GATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCAT 2590
CAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAATGTTGGCTCGACGATATC 2660
CACATCTGCCAAGAGCTTAGGTATCCGATCCTTCCGGCTTCTTTTTAGAAATTATATTATTTCAGAATCA 2730
TCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAACCTTCTAGACCAC 2800
      2810      2820      2830      2840      2850      2860      2870
AAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGAGGATCCCCGGGATTG 2870
GCCAAAGGACCCAAAGGTATGTTTCGAATGATACTAACATAACATAGAACATTTTCAGGAGGACCCTTGG 2940
AGGGTACCGGTAGAAAAAATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAAT 3010
TAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAA 3080
ACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGGTAAGTTTAAACATATATATA 3150
      3160      3170      3180      3190      3200      3210      3220
CTAACTAACCCTGATTATTTAAATTTTCAGCCAACACTTGTCACTACTTTCTGTTATGGTGTTCAATGCT 3220
TCTCGAGATACCCAGATCATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACA 3290
GGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTAAGTTTAAACAGTTCGGTACTAAC 3360
TAACCATACATATTTAAATTTTCAGGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCG 3430
AGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTC 3500
      3510      3620      3630      3540      3550      3560      3570
ACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTGTAAGTTTAAACTTGGACTTA 3570
CTAACTAACGGATTATATTTAAATTTTCAGAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTC 3640
AACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTA 3710
CCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTT 3780
GTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAGCATTCGTAGAATTCCAACTGAG 3850
      3860      3870      3880      3890      3900      3910      3920
CGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGAGTAAGTT 3920
TAAACTGAGTTCTACTAACTAACGAGTAATATTTAAATTTTCAGCATCTCGCGCCCGTGCCTCTGACTTC 3990
TAAGTCCAATTAGTGTTCAAGATCGGTAGATGGTGTTTGTGGTGTGGTGGGACGGGGTATTTTTGTTAT 4060
TATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTG 4130
TGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAAT 4200
      4210      4220      4230      4240      4250      4260      4270
AATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAAT 4270
TTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGAC 4340
CGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTT 4410
GGAGTATTTTTGGAATTTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGG 4480
GGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGA 4550
      4560      4570      4580      4690      4600      4610      4620
GCACGATGCAAGAAACATCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAA 4620
TTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACA 4690
TAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACC 4760
TCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCG 4830
TCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTA 4900
```

Monday, 1 December 1997 14:13 Page 3
fig pNP10

```
      4910      4920      4930      4940      4950      4960      4970
CTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGC 4970
CTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAG 5040
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTA 5110
TCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA 5180
CATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGC 5250
      5260      5270      5280      5290      5300      5310      5320
TGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAG 5320
CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTA 5390
TGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGA 5460
ATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG 5530
CAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG 5600
      5610      5620      5630      5640      5650      5660      5670
GAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA 5670
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT 5740
ATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT 5810
GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGC 5880
GTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC 5950
      5960      5970      5980      5990      6000      6010      6020
GACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAG 6020
CATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTA 6090
AAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCA 6160
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC 6230
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT 6300
      6310      6320      6330      6340      6350      6360      6370
TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG 6370
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT 6440
GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT 6510
CGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT 6580
ACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC 6650
      6660      6670      6680      6690      6700      6710      6720
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCG 6720
GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA 6790
CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG 6860
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAA 6930
CGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGC 7000
      7010      7020      7030      7040      7050      7060      7070
GCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGCTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAAC 7070
GCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT 7140
TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTGTA 7210
AGTTTAAACATGATCTTACTAACTAACTATTCTCATTTAAATTTTCAGAGCTTAAAAATGGCTGAAATCA 7280
CTCACAACGATGGATACGCTAACAACTTGGAAATGAAAT 7319
```

Figure 57

Monday, 1 December 1997 14:12 Page 1
ig pCB501

```
        10        20        30        40        50        60        70
ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCT  70
CGAGGATCAGAAGAAATTGGACCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAA  140
TTTTGAGTTTACGACTACAAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCTATTCTGTATGACT  210
AGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAAAGGAACATTATTTACTTTGCTTATTTGCCCTAAC  280
TTTGATTTAGTTTTTCGATCAACTAGATCTTACAAAACTTGCAATACAATTCCATTTTCAGATTACCCTC  350
       360       370       380       390       400       410       420
GCCACGTGTCGCCACGTCAGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACA  420
TCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGT  490
ATAAAATGCCCAGAAAAGAAGATGATAAAAATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTCTAA  560
TATGACGGCTTATATCTCAATTTTCTTGAGTTTTATCAAAAAATTTTCCACTATACAAATGTAGAAAAGT  630
ATTTTGCACAAATTTTGTCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAA  700
       710       720       730       740       750       760       770
AGTGGAAGGAGCGCAAGTCTATACTGGAAATAATGATCTGAAACAAATTTGTGCTATTCTCAAATGTTTA  770
AGACATGTTTTGAAGATTTTTTCAAATTCGCACTAGTTTCAGAACCTTCCTTTTTGTATGAAAAAGTAAA  840
AAAAAACTATTTCAAACCCTCACCGCCACCATGTTTCAACTCTTAATTTTTATAAAATTTTGCAATTTAC  910
AAATCGCCTCCCCTTGCCCGAAAAGTGCCCACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATT  980
CATGTGAGAAAACACAGAAGAGGCTAACTAAATTGACAGGGACAGGTTGTCCCTCTTCTCCCTCCTTCTC  1050
      1060      1070      1080      1090      1100      1110      1120
CCGCCTCCTCCTCCCGTTCCATCTCCAACAACAACAATTTTCCAATTTCGTTGTCCATTTTGCTTATAAA  1120
CATTTGTGTGTGGAAGGAAACTACACGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACTC  1190
TTTCGGAAATTGATGAATAAAGATACACCGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTT  1260
TTTCATCGTCTCAACTTGCGCATGAGTCTTCCCCCGCTCTCATCACTGACAATTAATGTCGGGTTTTATG  1330
CGCTCTTTCCTATTCCGCCACTCATTCTGGGTTACCACAAACTGGAATACATTTTACTACTATTCAAGCC  1400
      1410      1420      1430      1440      1450      1460      1470
ATTTATTTTGATATTTAATTTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACG  1470
ATATTTTCGATTTTAAAAAATCTGAAAAGTTTCAAAAAATCAATAAATATTCCCTAACAAATTGTATGGC  1540
TAAAATTTTATTTCTACTGTTGACAATATCTTTATATGTATCACTGTTTTCCATCTCAAAACCTTGAATC  1610
CCCCAAGTTATAGGAAGCTCCGTGTCACATTTCCCATGCTATGAATCGCTACTCAGCACATATCCAAAAA  1680
TTAAGCTAGACGGTTGATAATTATTGGGCACGCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAAGC  1750
      1760      1770      1780      1790      1800      1810      1820
ACAGATTATCTATCAAATTCAATCTTTGAACATTCACCCAGTTCGTACAATTTTCCATGCTTTTTGGCCC  1820
ATTAAAAAACTTTCTCACCTCTTCATCCATCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACTCT  1890
ACTTTTTAAGAGAAGGAGATACTGAGCCACATGGCGTGTGACCCTTTTCATCTCGTCCGTTCGGTCTCAA  1960
ATTCACGCTCATACTAACTCTTCAAATAGCCATAGACCTCCTTGTTTTCTTCTTCGTTTTGACTCGCGCC  2030
TATTTTTTGTGGCTGCCTGAAAGCCGGGAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA  2100
      2110      2120      2130      2140      2150      2160      2170
AAAAACGAGGCAATTTAAAAATATTAAAATTAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAA  2170
ACAAAACAAATAGGAACCGCCAGATCAAAATTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGG  2240
CTGAACAGAAAACTGAAGTGCCTGCATAAATCTAGTGTAACGTTTAGATTGAACTCGGAAATCCTAAGCC  2310
TGAACTATAGCCTTATTCTAGATCTTAGTTGCGCATAAGCTCAAGCCCAAGCAGAAATGACTTCCATTTA  2380
GTTTAAGCCTAGATTGACTTGCTTGCTTCAGTCTAATCCAGACTAGATTTCCAAGAGAGTTTTTCAATTTT  2450
```

Monday, 1 December 1997 14:12 Page 2
flg pCB501

```
      2460      2470      2480      2490      2500      2510      2520
AAATGTTTCCAGTTTCTTGTTACTTAAAATCTTAATGCCCTGTGATGCGTAAAATCGTTATCCCTTTCTC 2520
TCACACTTTCAATTACAGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTC 2590
ATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCCTTCGAGTGGCAATAATAAT 2660
GTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTATCCGATCCTTCCGGCTTCTTTTTAGAAATT 2730
ATATTATTTCAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800
      2810      2820      2830      2840      2850      2860      2870
AAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTA 2870
GCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTGCTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCG 2940
ATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCATCTTCCTC 3010
ATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCT 3080
GCCCCAGTGAAAAGTGGCCTGAAGCCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTT 3150
      3160      3170      3180      3190      3200      3210      3220
GTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAGACTCGAAACGATGCTC 3220
AAAGAGCAGTGAAGAAGAGTCCGGATACGCTGCATTCAACAGCACGTCGCCAACGTCATCATCGACGGAA 3290
GCTTCCCTAAGCATGCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACG 3360
ATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAA 3430
TATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAA 3500
      3510      3520      3530      3540      3550      3560      3570
GATCCACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCAC 3570
ATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACCAGAACCTGAAAAGCTCCAATCAATGAGCAT 3640
CGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCAATCCGA 3710
CAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATG 3780
AGCAGTCGTCCGCGTCTGAAGACTCCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAAC 3850
      3860      3870      3880      3890      3900      3910      3920
TTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATCCAGCGGCTACACCTCT 3920
GACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAG 3990
CACAGAACGGCTATCCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGA 4060
GCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGGCAACAGTCGCCTCCAAACATAGCGAC 4130
TATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTCGGTCCTCCACAT 4200
      4210      4220      4230      4240      4250      4260      4270
CAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCA 4270
ACGTGGCGCCGCTGCCACCTCAACCTTCGGACAACATTCGCTAAGATCCCCGGGATACTCATCCTATTCT 4340
CCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTATCCACTCACAGACTAGTCGACGACCTTCTT 4410
CACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATC 4480
CACCGAGCACAGAATGGCGGCTCTCTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCT 4550
      4560      4570      4580      4590      4600      4610      4620
TCAGGATCCTACTCGGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGAGAGACGTTCCAACTGC 4620
ACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGC 4690
TAGCACGACAGCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTCACTTG 4760
GAGTGTTACAAGAACACTGTCGACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTGTTTGATCTTT 4830
TTGAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACG 4900
```

Monday, 1 December 1997 14:12 Page 3

flg pCB501

```
       4910      4920      4930      4940      4950      4960      4970
ATTCAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAAC  4970
GAAGGCGCTGGTGAGCTTCTTCGTCAACCATCTCTGCAATCAGTTGCATCCCATCGATCATCGATGTCAT  5040
CGTCGTCGAAAAGCAGCAAGCAGGACAAGATCAGCTTGAGCTCGTTTGGCAAGAACAAGAAGAGCTGGAT  5110
CCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATT  5180
TCCGGATCTCAAGGAACTCTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATA  5250
       5260      5270      5280      5290      5300      5310      5320
GTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTTGATGTTCTGAGGGAGACAGT  5320
GAACAAGTTGAAAACGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACT  5390
CGTGCTTCTTCCCGCGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCA  5460
GTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGA  5530
CATCGCTGGAGAAATCAGTTCGATCGTTAACGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGC  5600
       5610      5620      5630      5640      5650      5660      5670
AAGGTCAGTGGAAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTT  5670
CTAAAATGGACCCAGCCTCTACCCTCGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGT  5740
GAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTC  5810
TCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGCTGTTCGAGACGCTGATCCCCAAGCCGATGA  5880
TGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAA  5950
       5960      5970      5980      5990      6000      6010      6020
GACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATC  6020
GTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACC  6090
AGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGG  6160
CTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACC  6230
ACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCA  6300
       6310      6320      6330      6340      6350      6360      6370
ACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGA  6370
CATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCAC  6440
AGCTTCCTTGAGAAGGCACAGGCAGCTCAG[illegible]CTTCCTCATCGGGGGTTGGTTC[illegible]TGTGTGGTGTGGCATTG  6510
GCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGG  6580
AGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGAC  6650
       6660      6670      6680      6690      6700      6710      6720
ACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGCCC  6720
CTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGA  6790
TCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACC  6860
ATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAAC  6930
GCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGC  7000
       7010      7020      7030      7040      7050      7060      7070
AGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTC  7070
CTAGGAACGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAAT  7140
GACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGCGCTTT  7210
CTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGG  7280
CTGAACTTGCTCAAAAGAAGCCGAATTCCAGCACACTGGCCTCCCCATGGTATTGATATCTGAGCTCCGC  7350
```

Monday, 1 December 1997 14:12
fig pCB501 Page 4

```
     7360      7370      7380      7390      7400      7410      7420
ATCGGCCGCTGTCATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAAC 7420
ATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTTCTTA 7490
ATTTCTTTGTTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATT 7560
AATTCGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCT 7630
ACACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAA 7700
     7710      7720      7730      7740      7750      7760      7770
AAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCA 7770
CGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTC 7840
AATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCA 7910
AGAGTTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGG 7980
AAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCT 8050
     8060      8070      8080      8090      8100      8110      8120
ATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGG 8120
CCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGG 8190
AGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGT 8260
TGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGG 8330
TGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGC 8400
     8410      8420      8430      8440      8450      8460      8470
CTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATG 8470
TGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACC 8540
CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATT 8610
CCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTG 8680
AAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCCCTAAGATCCTTGAGAGTTT 8750
     8760      8770      8780      8790      8800      8810      8820
TCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT 8820
ATTGACGCCGGGCAAGAGCAACTCGGTCCCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC 8890
CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAG 8960
TGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC 9030
AACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACG 9100
     9110      9120      9130      9140      9150      9160      9170
AGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTAC 9170
TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCG 9240
GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG 9310
CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTAT 9380
GGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA 9450
     9460      9470      9480      9490      9500      9510      9520
GTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCC 9520
TTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGA 9590
AAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCA 9660
CCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCA 9730
GCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGT 9800
```

Monday, 1 December 1997 14:12 Page 5

fig pCB501

```
        9810      9820      9830      9840      9850      9860      9870
AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGT 9870
CTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGT 9940
GCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAG 10010
CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGC 10080
ACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTG 10150
       10160     10170     10180     10190     10200     10210     10220
AGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT 10220
ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT 10290
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAG 10360
TGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATG 10430
CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC 10500
       10510     10520     10530     10540     10550     10560     10570
ACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGAT 10570
AACAATTTCACACAGGAAACAGCT 10594
```

Friday, 28 November 1997 11:36 Fig 58 Page
pTB115

Created: Friday, 28 November 1997 11:02

```
         10        20        30        40        50        60        70
ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTctagAAATGTAAACCTGTCATTTCTGTG  70
TATTTCAGCACCGCAGAGAGCACCATAAACAGCTACAACAAGAGCACGAACGTCGGCTCCAACAGAAAAA  140
GCTGCAAAACGCTGGTACTCACCACATGCTCTGATAATTGCCATTTCCTTCGATTTCTGACTTTCAATTG  210
TGATATgGATAACCTAAAATCTGCCATTCAAAATGAAAACTCTTAAAGTTTAAGAGGCTTTCAACGACCC  280
CGCGACTTTCCCACCTCATCGTTTTTCCCTTCTATCTTATTcTAATTTTTTTGTATTTTGTGTCAGTTTGC  350
        360       370       380       390       400       410       420
TATATCCTCATCACCTTTTTTTACCGTTTAATCTCTCTATAGTATTGTTCGGGTGGTTTCAAGATGAAACG  420
TTTTGTGTGTAGCATTTTTGAAACACAGcGxGAAAAATGCAGAAACTATCATCCACAATGCATTTAAGTGC  490
ACTCATTGTGATGACTCCTAACCCCCTcGCCCCACCATTTGTCTCTTCACAAATTCATGGCACAATTCAA  560
AAATGGCTTTTTGTGTTGGATCTTGATCTCGTCTCTTTCCTTCTTCTTTCGTTTCTCTCGGACCACTCAT  630
TATTCGAATGCCATGCATTTGTTGATGATGCGCACGATGGTCGCACACACTACAACAGATATGATGGTGG  700
        710       720       730       740       750       760       770
CTGTGCGTGGTGCAGAGCTCATGAATAAGTGATGGGGCACCGAGATGTGACTCCTCCCATTATTCCTCTT  770
GAGTCGCATCTTTTGTCTTGGTCCGTTGTCAGTCGAGCCTCAAGGCAGCCGTTCCGCCAGTGATCGCCTG  840
CTAAGAAGTTGTAGGTGTAGAGGAAGAAACGTCGGGTCCAAATTTCAAGcaacatccgcgagatcgtgaa  910
tcccgGGGATTGGCCAAAGGACCCAAACGTATGTTTCGAATGATACTAACATAACATAGAACATTTTCAG  980
GAGGACCCTTGGctagaactagtggatccgagctctcccatATGACGACGTCAAATGTAGAATTGATACC  1050
        1060      1070      1080      1090      1100      1110      1120
AATCTACACGGATTGGGCCAATCGGCACCTTTCGAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCC  1120
AATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAATTCTCGC  1190
CTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCT  1260
GAAAAATCTGGGTCTCGACTGCTCGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTT  1330
CTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAAAAAAGATCAGAAGAAAT  1400
        1410      1420      1430      1440      1450      1460      1470
TGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTC  1470
AGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCA  1540
CAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCAC  1610
CCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAA  1680
TAATGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCG  1750
        1760      1770      1780      1790      1800      1810      1820
AATCTAAACCGACCTACCTCCCAACTCCAAAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTA  1820
CAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTGCTTCTGT  1890
GAAGACTATTGGAGCAAAACAACAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAG  1960
TTATTCAGTAGCAAAAACCCATCTTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGC  2030
CTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAGCCGCCGACCAGTAAGCT  2100
        2110      2120      2130      2140      2150      2160      2170
GGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATC  2170
ATATCTCAACAAGACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACA  2240
GCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCATGCATTCCACATCTTCCAAGAGTTCAAC  2310
GTCAGACGAAAAGTCTCCGTCATCACACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAG  2380
CCGATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAG  2450
```

Friday, 28 November 1997 11:36 Page
pTB115

```
     2460      2470      2480      2490      2500      2510      2520
TGAAAGCAGTGAAAAGCACAGCGAAAAAAGATCCACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAAC  2520
AATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACCA  2590
GAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAG  2660
TTGTTCCACTTAAAATGACTTCAATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAAT  2730
CACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACTCCATTGTGGCTCATGCG  2800
     2810      2820      2830      2840      2850      2860      2870
TCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATA  2870
AGACATCAGAATCCAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCT  2940
GAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTATCCTGACAACTTCGAAGACAGTTCCTCC  3010
TTGTCGTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACA  3080
TGGCAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAA  3150
     3160      3170      3180      3190      3200      3210      3220
GCCCCGAGTCCCCAGTCGGTCCTCCACATCAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTAC  3220
AAACTTCTGTCCCAGTGCCGAACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGGACAACATTCGC  3290
TAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTAT  3360
GCACTCACAGACTAGTCGACGACCTTCTTCACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGAT  3430
CGTAAATGCCACCTTCAAGAGTTCACATCCACCGAGCACAGAATGGCGGCTCTCTTGAGCCCGAGACGGG  3500
     3610      3620      3630      3640      3650      3660      3670
TGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCCTACTCGGCGCGTTCCCGAGGTGGAAGCTCTAC  3570
TGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCCAAA  3640
AGTGAGATGGGATCCCAACTATCACTGGCTAGCACGACAGCATATGGATCTCTCAATGAGAAGTACGAAC  3710
ATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGTCGACTCACTAACCAAGAAACA  3780
GGAGAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAACTCACTCAACACATTGATCGA  3850
     3860      3870      3880      3890      3900      3910      3920
TCCAACTTGAAGCCTGAAGAGGCAATACGATTCAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATC  3920
ATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCTTCGTCAACCATCTCTGGAATC  3990
AGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGC  4060
TCGTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGA  4130
ACTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAACTCTTGACAACATTGATGTGATTGA  4200
     4210      4220      4230      4240      4250      4260      4270
GTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCC  4270
CGCGAAGTTGATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAG  4340
TGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCGCGCCTCAATTCCAGTTATCTACGACGA  4410
TGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGC  4480
AACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAG  4550
     4560      4570      4580      4590      4600      4610      4620
AGATAATCGTAGGATATCTTGCCATGTCAACCAGTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCT  4620
AGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATCAACTTGGAATCGATGCTCGTGATTCT  4690
ATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCATC  4760
CAACTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCATGCACGGTGCCGCACAGAGTCGCGTAGA  4830
CAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACTCGTCAAGTCAATTTTGACAGAG  4900
```

Friday, 28 November 1997 11:37 Page
pTB115

```
        4910      4920      4930      4940      4950      4960      4970
AGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTATG 4970
TATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAAGA 5040
ATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGCAAAGAATCATGCATCGTAATTCTAGAT 5110
AATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGTCCCACTTCAAAACAACGAAG 5180
GTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAAAT 5250
        5260      5270      5280      5290      5300      5310      5320
GTCAGTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGACGACGGCCGGTAGAGGATGAG 5320
TATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCTCTTCAGG 5390
CCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACTG 5460
TCCTCTAACTGTCGATGGATCCCGTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATTTG 5530
GAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCACTTCCTTCGAGGATCCCACCGACATCG 5600
        5610      5620      5630      5640      5650      5660      5670
TCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAAGA 5670
CCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCCTCGAGTCGTTGATCCAATTGCAT 5740
GCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATCTTCTCTCGCCTCTCCCCCGCT 5810
TTCCTTATCTTCGTACCGGTACCATGGTATTGATATCTGAGCTCCGCATCGGCCGCTGTCATCAGATCGC 5880
CATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCT 5950
        5960      5970      5980      5990      6000      6010      6020
GTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTTAGCTTCTT 6020
TTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAA 6090
AAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTC 6160
TTATGTTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTA 6230
TCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCA 6300
        6310      6320      6330      6340      6350      6360      6370
AATCATGCTCATCGTGAAAAAGTTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAAT 6370
TAATTTTCCTGCTTTTCCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTT 6440
CTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTTGAGGCTC 6510
AGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGAC 6580
AGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGC 6650
        6660      6670      6680      6690      6700      6710      6720
GCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA 6720
GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTA 6790
ACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGT 6860
AAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCAT 6930
GATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTA 7000
        7010      7020      7030      7040      7050      7060      7070
TTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT 7070
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCT 7140
TCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTG 7210
GGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAA 7280
TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACT 7350
```

Friday, 28 November 1997 11:37 Page
pTB115

7360 7370 7380 7390 7400 7410 7420

CGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACG 7420
GATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTAC 7490
TTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCG 7560
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTA 7630
GCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAA 7700

7710 7720 7730 7740 7750 7760 7770

TAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTAT 7770
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG 7840
CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG 7910
CTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGAT 7980
TGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA 8050

8060 8070 8080 8090 8100 8110 8120

ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG 8120
ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTT 8190
GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACT 8260
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTC 8330
TGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACG 8400

8410 8420 8430 8440 8450 8460 8470

ATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGA 8470
ACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAA 8540
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA 8610
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG 8680
TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC 8750

8760 8770 8780 8790 8800 8810 8820

CTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG 8820
AGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGC 8890
CCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCG 8960
ACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTT 9030
ACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGC 9100

9110 9120 9130 9140 9150 9160 9170

T 9101

Friday, 28 November 1997 11:00 Page
pPD95-75

Created: Friday, 28 November 1997 10:58

```
        10        20        30        40        50        60        70
AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGATTGGCCAAAGGACCCAAAGgtatgtttc  70
gaatgatactaacataacatagaacattttcagGAGGACCCTTGGAGGGTACCGGTAGAAAAAATGAGTA  140
AAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA  210
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACT  280
ACTGGAAAACTACCTGTTCCATGGgtaagtttaaacatatatatactaactaaccctgattatttaaatt  350
       360       370       380       390       400       410       420
ttcagCCAACACTTGTCACTACTTTCTgTTATGGTGTTCAATGCTTcTCgAGATACCCAGATCATATCAA  420
ACgGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGAT  490
GACGGGAACTACAAGACACgtaagtttaaacagttcggtactaactaaccatacatatttaaattttcag  560
GTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGA  630
AGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGAC  700
       710       720       730       740       750       760       770
AAACAAAAGAATGGAATCAAAGTTgtaagtttaaacttggacttactaactaacggattatatttaaatt  770
ttcagAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAA  840
TACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCG  910
AAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATG  980
GCATGGATGAACTATACAAATAGCATTCGTAGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTT  1050
      1060      1070      1080      1090      1100      1110      1120
GTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGAgtaagtttaaactgagttctactaactaacga  1120
gtaatatttaaattttcagCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCC  1190
CTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTTCTTAATTT  1260
CTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATT  1330
CGTAATAAAAAGTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACAC  1400
      1410      1420      1430      1440      1450      1460      1470
AATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAA  1470
CCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCACGTC  1540
TGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATC  1610
AAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAG  1680
TTTCGAGGACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAGA  1750
      1760      1770      1780      1790      1800      1810      1820
AGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATGG  1820
GGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCCTACTAGTCGGCCCT  1890
ACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC  1960
GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC  2030
GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTG  2100
      2110      2120      2130      2140      2150      2160      2170
AAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTAT  2170
TTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCG  2240
CGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA  2310
TAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCT  2380
TTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGA  2450
```

Friday, 28 November 1997 11:00 Page
pPD95-75

```
     2460      2470      2480      2490      2500      2510      2520
TCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGC 2520
CCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTG 2590
ACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGT 2660
CACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGAT 2730
AACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACA 2800
     2810      2820      2830      2840      2850      2860      2870
TGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCG 2870
TGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTA 2940
GCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCC 3010
TTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGC 3080
ACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGAT 3150
     3160      3170      3180      3190      3200      3210      3220
GAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTT 3220
ACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTT 3290
TGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAG 3360
ATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC 3430
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAG 3500
     3510      3520      3530      3540      3550      3560      3570
AGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA 3570
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTA 3640
CCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCAC 3710
ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCC 3780
ACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA 3850
     3860      3870      3880      3890      3900      3910      3920
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG 3920
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGG 3990
TTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACC 4060
GTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAG 4130
CGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGC 4200
     4210      4220      4230      4240      4250      4260      4270
TGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTC 4270
ATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACA 4340
ATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTgtaagtttaaacatgatcttactaacta 4410
actattctcattaaattttcagAGCTTAAAAATGGCTGAAATCACTCACAACGATGGATACGCTAACAA 4480
CTTGGAAATGAAAT 4494
```

Friday, 28 November 1997 13:10
pBS KS/X16

Page

Created: Friday, 28 November 1997 12:02

```
          10        20        30        40        50        60        70
CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAC 70
CAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTC 140
CAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCA 210
GGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTA 280
AATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG 360
         360       370       380       390       400       410       420
AAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCAC 420
CACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGG 490
GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGAT 560
TAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATA 630
CGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGG 700
         710       720       730       740       750       760       770
CTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAGAAATTGGAGCAACTAC 770
CCACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAATTTTGAGTTTACGACTACAAAAATGTGTT 840
CTTTAATAACTATCTTCGACTTGAGTCTATTCTGTATCACTAGTTGTTGAGTGATTTTTCATTGAGAAAA 910
TATTAAAAGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTTAGTTTTTCGATCAACTAGATC 980
TTACAAAACTTGCAATACAATTCCATTTTCAGATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTT 1050
        1060      1070      1080      1090      1100      1110      1120
CAGCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAAT 1120
ATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGTATAAAATGCCCAGAAAAGAAGATGATAAA 1190
AATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTCTAATATGACGGCTTATATCTCAATTTTCTTGA 1260
GTTTTATCAAAAAATTTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAATTTTGTCAGTTGACAGC 1330
TTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAAAGTGGAAGGAGCGCAAGTCTATACTGGAA 1400
        1410      1420      1430      1440      1450      1460      1470
ATAATGATCTGAAACAAATTTGTGCTATTCTCAAATGTTTAAGACATGTTTTGAAGATTTTTTCAAATTC 1470
GCACTAGTTTCAGAACCTTCCTTTTTGTATGAAAAAGTAAAAAAAAACTATTTCAAACCCTCACCGCCAC 1540
CATGTTTCAACTCTTAATTTTTATAAAATTTTGCAATTTACAAATCGCCTCCCCTTGCCCGAAAAGTGCC 1610
CACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATTGATGTGAGAAAACACAGAAGAGGCTAACT 1680
AAATTGACAGGGACAGGTTGTCCCTCTTCTCCCTCCTTCTCCCGCCTCCTCCTCCCGTTCCATCTCCAAC 1750
        1760      1770      1780      1790      1800      1810      1820
AACAACAATTTTCCAATTTCGTTGTCCATTTTGCTTATAAACATTTGTGTGTGGAAGGAAACTACACGGG 1820
GAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACTCTTTCGGAAATTGATCAATAAAGATAGAGC 1890
CGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTTTTTCATCGTCTCAACTTGCGCATCAGTCT 1960
TCCCCCGCTCTCATCACTGACAATTAATGTCGGGTTTTATGCGCTCTTTCCTATTCCGCCACTCATTCTG 2030
GGTTACCACAAACTGGAATACATTTTACTACTATTCAAGCCATTTATTTTGATATTTAATTTTGTGCAAT 2100
        2110      2120      2130      2140      2150      2160      2170
TAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACGATATTTTCGATTTTCAAAAAATCTGAAAAG 2170
TTTCAAAAAATCAATAAATATTCCCTAACAAATTGTATGGCTAAAATTTTATTTCTACTGTTTGACAATAT 2240
CTTTATATGTATCACTGTTTTCCATCTCAAAACCTTGAATCCCCCAAGTTATAGGAAGCTCCGTGTCACA 2310
TTTCCCATGCTATGAATGGGTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGATAATTATTGGGC 2380
ACGCCTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAAGCACAGATTATCTATCAAATTCAATCTTTGA 2450
```

Friday, 28 November 1997 12:03
pBS KS/X16

Page

```
    2460      2470      2480      2490      2500      2510      2520
ACATTCAGCCAGTTCGTACAATTTTCCATGCTTTTTGGCCCATTAAAAAACTTTCTCACCTCTTCATCCA 2520
TCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACTCTACTTTTTAAGAGAAGGAGATACTGAGCCA 2590
CATGGCGTGTGACCCTTTTCATCTCGTCCGTTCGGTCTCAAATTCACGCTCATACTAACTCTTCAAATAG 2660
CCATAGACCTCCTTGTTTTCTTCTTCGTTTTGACTCGCGCCTATTTTTTGTGGCTGCCTGAAAGCCGGGA 2730
AAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATAAAAAACGAGGCAATTTAAAAATATTAAAA 2800
    2810      2820      2830      2840      2850      2860      2870
TTAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAAACAAAACAAATAGGAACCGCCAGATCAAA 2870
ATTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGGCTGAACAGAAAACTGAAGTGCCTGCATAA 2940
ATCTAGTGTAACGTTTAGATTGAACTCGGAAATCCTAAGCCTGAACTATAGCCTTATTCTAGATCTTAGT 3010
TGCGCATAAGCTCAAGCCCAAGCAGAAATGACTTGCATTTAGTTTAAGCCTAGATTGACTTGCTTGCTTC 3080
AGTCTAATCCAGACTAGATTTCCAAGAGAGTTTTCAATTTTAAATGTTTCCAGTTTCTTGTTACTTAAAA 3150
    3160      3170      3180      3190      3200      3210      3220
TCTTAATGCCCTGTGATGCGTAAAATCGTTATCCCTTTCTCTCACACTTTCAATTACAGATTCATCAAAG 3220
ATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAA 3290
ATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAATGTTGGCTCGACGATATCCACATCTGCGAA 3360
GAGCTTAGGTATCCGATCCTTCCGGCTTCTTTTTAGAAATTATATTATTTCAGAATCATCATCAACGTAC 3430
AGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAACCTTCTAGACCACAAACCCAGCTAG 3500
    3510      3520      3530      3540      3550      3560      3570
TTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAA 3570
ACTTGCTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATG 3640
CTGAAATTAAAGTTATTCAGTAGCAAAAACCCATCTTCCTCATCGAATAGCCCACAACCTACGAGAAAGG 3710
CGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAGCCGCC 3780
GACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGGTGAGTATTTGAAATCGGAAATTG 3850
    3860      3870      3880      3890      3900      3910      3920
GAAATGTATTTTTTAAAAACTGAAAATTCTACAAAATAAATTAAAAATAAGATTTTTTCCTCTGATAGTA 3920
TTGCATCCCACTATTTTACTTTGAAGATTTATATCTTGGTTCATATTGAAGATATCAGATATAGAAAAAG 3990
AAATAAAAAAATATTTTGACAGTTGATAATTTTTGTATAGGACCAAAGACAAGTGAGATATAAGCTGTCAA 4060
AGTTGATTTTCAAGAAATTTTAAAACCCTAGTTTTGCGAAGCTCTGGGCCTCATCTATATTTAGAACCCG 4130
ATTCGTACTTCTTCCGTTCCTTGACTCTACCAAAACCAAAACCAACCTACTAATAAAAATGATGAGACAA 4200
    4210      4220      4230      4240      4250      4260      4270
TTGGGAATTGTCTCCCATTTTTCTTCTTTCTCCTTGACACTCTTTCAGAATCTATGTCCCATCTTTTTGTC 4270
GTTGTGTCCCCCCATAAAGACTCTTCGCGGAAAAATGTTGCAACGGAAGTGATATTCCGAGCATTTTTCG 4340
ACGTCGAGGGCCGAAAAACACATCTGGCTGACAAAGAGTAAAGCAATTTCTCAGCTCTTCTTCGCCGGTT 4410
TTTCAATTCGTTTTTCAAAATGAGCTACTACAGAGTGAAAGAGCACAAATTGCAAACATTTTTGTGTGA 4480
GATGCACTTTTTGAAAAATTAACTTTACGTTTTCAGTTTCTAGTATTTATTTTTTTCATATAAATTAGAG 4550
    4560      4570      4580      4590      4600      4610      4620
CTTCTTAGACCTGCTATATTTTTTAAAACTTCCTACTGAAATATACGAGATTCTTTTGACTTTCCGGAAT 4620
TGTCTTATGGCTTCTATTATTTATGAGAAAACATTTTTTAAAAATTTTTTTGAAAAAAAAACTGTGCATC 4690
TTCGTTTTTTACATAGTAATTTCCAGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCA 4760
ACAAGACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCG 4830
CCAACGTCATCATCGACGGAAGGTTCCCTAAGCATGCATTCCACATCTTCCAAGGTTCGTTGTTTAGGAG 4900
```

Friday, 28 November 1997 12:03 Page
pBS KS/X16

```
       4910      4920      4930      4940      4950      4960      4970
AACTGTTTTTTTTTGTTTTCGTGACCTTCACATAGTCTCGGATGTTTATAAAAGTGAGGTCTCTGGACA 4970
CCTGCCATAAAATGTGAATCCGCCCATTTGTTGGTACAAAAAACTTTGCAGAGCACCTGCTTTATACATT 5040
TTTAGGATAAATGTCATACGGTATTTGTCAAACCCAAACTTTTTAAATTTTATTTTTCAGATCAAAAATTG 5110
ATGTTAAAAGTTTAAGATATTTACGAAAAAATGTTTTACTTAAAACTTTTTTATATGGATAAAATTTTAG 5180
AACATTCAGATAGGAGTTCCGTCCCTAAAACTTTTGTGGTCGCCGAGAAGCTTTCGAATTAATAATCTCA 5250
       5260      5270      5280      5290      5300      5310      5320
TTTTATAAGTTCGAAACTAATTTTTTGTGCAACTGAATTTTTAGAGATGAAACTTTAAGTTTCAATTTAT 5320
CCCATTTTGAAACCGTCCCCTTCTATAAAACTTCAAAATTTTCAGAGTTCAACGTCAGACGAAAAGTCTCC 5390
GTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCG 5460
GTTTCTCCAAATATTATCAACAAGCCTGTTGAGGTGAGTATTTTTTGTTTCTGGGTAGAGGCTTCTTGTC 5530
AAGTTTGGCCTAAATTATATTAACTTGGTTTAGAGGCTGGCAAAGCCATTGATCAAGCATGGGCTAAACT 5600
       5610      5620      5630      5640      5650      5660      5670
GGGCCGGCCTGAACCATGTACATATTTCTTTGGCCCGAGTAGTTGCAATCTAAAGATTCGAAGCTGGCTTC 5670
AAAGTCGGACTAGGCAAAAGTGCAAAAATGGAAAATATCTTGAAATTCAATACGCTTTCCGTCTTTCCAT 5740
TCTTCCTTTTTTGTCGTCTTTTGTTGCAGATTTTCCCTTTTTTAGATTTTAAGATTTTGATACACTTTAA 5810
TGTCTGGCTTCGCCTTCCTAAGAGCCTTCCTATATTTTCGAAAAATAATCAATTTTTAGGAAAAACCAAC 5880
ACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATCCACCTCCAGCTGTTCCGCCACGTGACACC 5950
       5960      5970      5980      5990      6000      6010      6020
CAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGGTACGTTTATTTCTG 6020
AACTTTACTTATGTTTCGGTCGGTGACGTTTTTGTTGACCATGTGATGGGAAGTAATTTTTGGATTATTT 6090
AAAGTTTGTGCGGGAATAGTAAGGAGGAGTACAATTATTTTATTTGTAGAAGGTTCAGTAAAACTTTGAT 6160
TTTTCTGACCATAAGTTTTTTTTCTGAAAGTTGTTTAAAAAATTGAGTTAAAAAATAAAATAATACTCTA 6230
TAAAAATTTCTAAATTCTTGGGAATTTTTTTCAAAAATGTTTTTCCAAATATGTCTAATAGTAAGATTTG 6300
       6310      6320      6330      6340      6350      6360      6370
TTTGTGAATTTACAAAACATATTTTAAAATACATTTTAATTTATTCGATTTTTTCGTTCGCAGGAACGAC 6370
AAAAAATCAGAAAAAGCGAAATTTAATTCCAAAAAAAATATTTTGAAAACTCACAAATAAAGCTACTTTT 6440
CAAAAAATCAACAAAAAAAATATCAAAAGATTCATATTTTCAGAAATAGAGACATAGTCAAAAACTATCA 6510
AAAAATTCACTATTTTCCGGAAAAAAATTGAGAAAAATTCCAAAAATTGTAAAAAACAAAATTGAGAAAA 6580
ATTCCAGAAATTGAAAAAAAATTCCTTTGAGGAAAATTTAAAAATTTTAAATGTGTGATTTCTGAAACCA 6650
       6660      6670      6680      6690      6700      6710      6720
AGCATTTTCCGACTTTTCGGCGATTTTCAGACTTGGGCTATAAATTTTTGTCAAAACTAGGAATCTTAAA 6720
ATATTTGTATTTTTTCGAAGAATCCTCCTCAATCTCAAATTCATATTTTATAATTTCAGACCCCGTGATAT 6790
CTGAAAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACC 6860
TCTAAAATCAGTTGTTCCACTTAAAATGACTTCAATCCGACAACCACCAACGTACGATGTTCTTCTAAAA 6930
CAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGTCAGTGCACCCCCCACCTCCAATTATTATGACAAA 7000
       7010      7020      7030      7040      7050      7060      7070
TGACCATTTTGCAGGATATGAGCAGTCGTCCGCGTCTGAAGACTCCATTGTGGCTCATGCGTCCGCTCAG 7070
GTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAG 7140
GTAAATTTTGGAAACTTTGATTTTTTTGTTGAAAAATAGCTTCAAATTATAAATTTTAAAAAATCCCAG 7210
AAAAATGATGTTTGTCAAAGGAAAACTTTTGATTTTTTGGTTTCTGAACTGTTTCGTTTAAAGTTAACCA 7280
ACGTTGGAGCTCGTACCAAAAACTTTTTCTTTTGATAATTTTTGAATCTATAGTATTTTAATTTTTGGAA 7350
```

Friday, 28 November 1997 12:03
pBS KS/X16

Page

```
      7360      7370      7380      7390      7400      7410      7420
TTGTGAAAGTTCTCTTGAGATGTATTAAGTTTTAGGCATAGGCAGGTGTGTAGGCAGAAAGGTATGATGT 7420
AGGCAGATAGGCTTGATATTTACCAAGCCAATAAACAGTAAATAATATTTAAAAAAAAACACTGAATAAA 7490
TCAAAAGCTAATAATTATTGTTTATTGGACCTACCAACACCTTACATTTGCCTACATGCTTACCTATTCC 7560
TTGTTGTCTACATTTTGAACGTTAATCACTAATTCGGTGAATGAACACTTGTAGATTTTTAATTTCGACA 7630
GTAATTTTTGAGCACATTGGCGTTAGAATTTGAAAAAAAACCTTCGACAGTTGAATCCTCATAACTCTCA 7700
      7710      7720      7730      7740      7750      7760      7770
AAATATTTCAGAATCCAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAG 7770
CTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTATCCTGACAAGTGAGTTTTGGTAGTAG 7840
TAGTTGTAGTCCCTTGACACACATATGAACACATTCGCTGCTCGTTTCGGTGGTCAGGGAGCCATGGAGC 7910
AATTAATCCAGAAGGCTCAAAATTAATGAGCATCACTTGGTGATCGAGGAATCCCCGAAAGACGTTTGAT 7980
AGCATCTTCTTCTTTTGCATTCTTTTTCTCTCTCTGCTGGGAGTCCCTGTTACACAGACATCTATTCATG 8050
      8060      8070      8080      8090      8100      8110      8120
CGCGAAGTGCAATTTTGGTTCCTAAAGATGAGAGGAGGAGGAGGAGCTATGACTTATGAATGAGCATCGA 8120
GGAACCGCGTGAAATAGTTTGGAGCTTAGATTGCAAATTACAGGAATATTCCGGTGACCTCAGTCTACTT 8190
GGCATTTGGCGGCAGTTTGGTAAAACTTGAGCCAAATTTTGTTTGGGTCACGCGTAATTTTCAAACAGTG 8260
GCAAGAATTTGCTCAACAAGTCTTGCTGGCCCAAGTCTTTCTAGAGTCTGTACCAAGCTTGGTCCAATAC 8330
TTTTTGGGCCAAACTTTGGGAAGAAGTTGCCGCCAAATGCATTTTTTCAAAAATACTAGAACCAACATCT 8400
      8410      8420      8430      8440      8450      8460      8470
GGTTTAGCGCAAGTTGTCTCCTAGGAGGATATCCAAATGTTTATTTACTCTCTCTCTTTCTAACGAAGAT 8470
CTATCGGTGCATCTCATAATGGAACCGATGCGGTTTTGCGCGCTTCGAAGAGCATTTCTTTGCTTTTGTT 8540
GCTTTTTTGTTATGCCTCTGAATTTGAAACTAGTTATGAAATTTCAGATGTTTCAGTGGAGATGGCCAAG 8610
ATAGTTGGTATTACGAAAGTTTCATGAATTTTGAAATTCTGTAGTTCTGTGAGTCTGCATACTATAGCTC 8680
AAAAAGGTTCCGCTTGAGCCCTGCCTAAATTCAAAATTTCCTTTCAAGTCCTGTCCATTTAATTTTAGGTT 8750
      8760      8770      8780      8790      8800      8810      8820
ATGAAATAACGCGAACCTATCATTCTTCTTCGTCTCAATTTCTTTTTCATGTCACCCCGATCCTGTGGGT 8820
TTCAATCTTCTTTCAATTCACCCTAGATTTGGCGCACAAGCTCCCCTCTCTCTCTAATGCAGCTGTCCTC 8890
CCGCTTCCATTTTTGGCGACTTGCTTCCTTCCCTTCCCTTCCGTCGCCGGTTTCCTTCCGATTGCCTTGT 8960
GTTCCTATAATTAATTCATTAAGCCGAGAACAAACGAACGGGGTTTCTTTTTCTCCTCTTACCATCTTTT 9030
TGGCGGAATTGAATTCGTTTTGTGATAGTGAAGTGTTTGTCAAGAATTTTTGGTTTTTGGTAGCTTGCCA 9100
      9110      9120      9130      9140      9150      9160      9170
AATTATTGTTAGTATGCACCATGTTGCTCAGCATTGTGTTCAATTATCTTGGTTTTATTTCAAAATTTGTT 9170
TCATAAAAACTTTTGTAGGCAGGTTTTCGTGGCGGGACCCACTTTTTCAAACTCTAGCCAAAAATGTGTG 9240
TCAATTCGGATCTCAAATTATCTCAACTATAACCTTAGAACGTTTTTTTCAAAAATTTCCGAAATCATAT 9310
TTTTCGAAATTCCAATTTTTTTAAAGATTAAACTGCCAAATTTGAATTCCATGCGCCTTCCATTTGACTC 9380
CCATCATGACTACGTTCCGTGGGTCTCGCCACAAAAATTACAGTTCCTCGGAAGGTTTTTGTGGCGGGGC 9450
      9460      9470      9480      9490      9500      9510      9520
CTAGCAGAGCCCATAGGTGCCAAAGCTCCCGATGAGCTGATAAAAAATGTACTTCTAAGGAATATTATGC 9520
AAAAATAATTTGGAGCTTTTTTTACTAGAAACAGTTTAAGAAAGAAAAAAGGTTTTTTTTTTAATTTAAAAA 9590
TTCAAATTTTGAATATTAAAGCCAAATTTGAGTTGATCACTTCGAGAAAATTCAAAGATTGAAAGCCTAG 9660
ATTTTTGATGCAATTTATCTGGAAAGCGAACTTCTAAATTAGAAAAACTTATAAAAAAATCTAAATGTTT 9730
GAAATTTTTTATTTGAAATCCTAGTTGACTTTTTTTGATTTTTCTATTTGTTTCCAAGGTATAAGTCTTT 9800
```

Friday, 28 November 1997 12:03 Page
pBS KS/X16

9810 9820 9830 9840 9850 9860 9870

GAAGTCGCCTCCTCAACCTCAAACCAGTGTGCCTCCATATTTGGAACACACACAAGCAAAAACCAATTGA 9870
TAGTATGTGTTCGAGTAGCCACTTGACAAGAAGAAACTTGCCGACACTGGTGGCTGGTCACCATTCTCCT 9940
CTCTTTGTCATTTGCATAATCTTTCTCCCTCTTCCTCATAAATAACTAAACTGTGTGTCCTGCGCGTCCC 10010
CCGCTCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGT 10080
AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGG 10150

10160 10170 10180 10190 10200 10210 10220

AAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTG 10220
CCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 10290
GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG 10360
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC 10430
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC 10500

10510 10520 10530 10540 10550 10560 10570

TCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA 10570
AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA 10640
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT 10710
CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT 10780
ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT 10850

10860 10870 10880 10890 10900 10910 10920

AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT 10920
ACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG 10990
CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGC 11060
AGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT 11130
CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATG 11200

11210 11220 11230 11240 11250 11260 11270

AAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG 11270
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA 11340
CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC 11410
AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCC 11480
TCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACG 11550

11560 11570 11580 11590 11600 11610 11620

TTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC 11620
CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG 11690
ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA 11760
CTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG 11830
TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA 11900

11910 11920 11930 11940 11950 11960 11970

AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA 11970
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG 12040
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA 12110
CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT 12180
GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC 12237

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence (fig 561)
Enzymes : 100 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

```
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA 100
  T  D  R  E  I  S  R  S  P  M  V  D  S  Q  Y  N  L  L  .  C  R  I  V  K  P  V  S  A  P  C  L  C  V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACTGGCTGTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGC 200
 G  G  R  .  V  V  R  E  Q  N  L  S  Y  N  K  A  R  L  D  R  Q  L  H  E  E  S  A  .  G  .  A  F  C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGGTCTATATGCGCAACTGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT 300
A  A  S  R  C  T  G  Q  I  Y  A  L  T  L  I  I  D  .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCA 400
 G  V  P  R  Y  I  T  Y  G  K  W  P  A  W  L  T  A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG 500
 N  A  N  R  D  F  P  L  T  S  M  G  G  L  F  T  V  N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A

CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT 600
P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V  H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTT 700
 R  Y  Y  H  G  D  A  V  L  A  V  H  Q  W  A  W  I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q

TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG
ACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTC 800
W  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .  Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG 900
                                                                            T7 promotor priming site
G  L  Y  K  Q  S  S  L  A  N  .  R  T  H  C  L  L  A  Y  R  N  .  Y  D  S  L  .  G  D  P  S  V  L  A

GTTTAAACTTAAGCTTACCATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG 1000
                                           ProBond binding domain
 F  K  L  K  L  T  M  G  G  S  H  H  H  H  H  H  G  M  A  S  M  T  G  G  Q  Q  M  G  R  D  L  Y  D
```

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 2

GATGACGATAAGGTACCTAAGGATCCAGTGTGGTGGAATTCTGCAGATATCGAATTCCTGCAGCCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCC
CTACTGCTATTCCATGGATTCCTAGGTCACACCACCTTAAGACGTCTATAGCTTAAGGACGTCGGGGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGG 100

insert pLM1

ORF pLM1

D D D K V P K D P V W W N S A D I E F L Q P L L F S Q M L D P E S

AGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGA
TCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCGGACGCTCCCAGGGTCCACTGAGTGTCGAGGGACCT 200

insert pLM1

ORF pLM1

Q R K R T V Q N V L D L R Q N L E E T M S S L R G S Q V T H S S L E

GATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCG
CTACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAGACAGTACCGCGATACCGGTCAGGTCAGGC 300

insert pLM1

ORF pLM1

M T C Y D S D D A N P R S V S S L S N R S S P L S W R Y G Q S S P

CGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTACATGCACGGCGAACGGGCCCACTACTCCCACA
GCCGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGT 400

insert pLM1

ORF pLM1

R L Q A G D A P S V G G S C R S E G T P A W Y M H G E R A H Y S H

CCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACAT
GGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTA 500

insert pLM1

ORF pLM1

T M P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y M

GAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCC
CTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTACTTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGG 600

insert pLM1

ORF pLM1

S D S D L M G K T M T E D D D I T T G W D E S S S I S S G L S D A

TCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGCAGGAACTCAACAATAGTGCTAC 1700
AGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACGAAGAGCGTCCTTGAGTTGTTATCACGATG

insert pLM1

ORF pLM1

S D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L

GCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCCTAAAAAACTGGAGTACGACAGTGGTAG 1800
CGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTCTTTCGGGGATTTTTTGACCTCATGCTGTCACCATC

insert pLM1

ORF pLM1

R T D S E K R S L A E S G L S V F S E S E E K A P K K L E Y D S G S

CCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCATCCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTG 1900
GGACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCCCACCTCTTGACTTTTTCGGGTAGTCGGAC

insert pLM1

ORF pLM1

L K M E P G T S K W R R E R P E S C D D S S K G G E L K K P I S L

GGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCACACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAAC 2000
CCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCGGGTCTCACGGGAGTTTCAGCGTCCGTTTG

insert pLM1

ORF pLM1

G H P G S L K K G K T P P V A V T S P I T H T A Q S A L K V A G K

CTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGATGCTGGTCGGGACCGCCTGAGTGATGCTAA 2100
GACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGGAGACTACGACCAGCCCTGGCGGACTCACTACGATT

insert pLM1

ORF pLM1

P E G K A T D K G K L A V K N T G L Q R S S S D A G R D R L S D A K

GAAGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTCCTGCCACAGGCACAGCCACTGTCATGCAAACTGGT 2200
CTTCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAGGACGGTGTCCGTGTCGGTGACAGTACGTTTGACCA

insert pLM1

ORF pLM1

K P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence


```
GGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGC
CCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCG 2300
                                         insert pLM1
                                         ORF pLM1
 G  S  A  T  L  S  K  I  Q  K  S  S  G  I  P  V  K  P  V  N  G  R  K  T  S  L  D  V  S  N  S  A  E

CAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGTTCTATGAGCGTGACCGGCGGGCGGGGTGG
GTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCAAGATACTCGCACTGGCCGCCCGCCCCACC 2400
                                         insert pLM1
                                         ORF pLM1
P  G  F  L  A  P  G  A  R  S  N  I  Q  Y  R  S  L  P  R  P  A  K  S  S  S  M  S  V  T  G  G  R  G  G

ACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGT
TGGAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGGTCTGACTTCCTCGGATGGTTCCATCGGTCA 2500
                                         insert pLM1
                                         ORF pLM1
 P  R  P  V  S  S  S  I  D  P  S  L  L  S  T  K  Q  G  G  L  T  P  S  R  L  K  E  P  T  K  V  A  S

GGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGA
CCCGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCGTCACCGGAACCTGAGTCTGTTGTAGAGGAACTTCT 2600
                                         insert pLM1
                                         ORF pLM1
 G  R  T  T  P  A  P  V  N  Q  T  D  R  E  K  E  K  A  K  A  K  A  V  A  L  D  S  D  N  I  S  L  H

GTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAA
CATAACCGAGGGGTCTCTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGTCTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTT 2700
                                         insert pLM1
                                         ORF pLM1
S  I  G  S  P  E  S  T  P  K  N  Q  A  S  H  P  T  A  T  K  L  A  E  L  P  P  T  P  L  R  A  T  A  H

GAGCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAAGGTC
CTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATGGTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAG 2800
                                         insert pLM1
                                         ORF pLM1
 S  F  V  K  P  P  S  L  A  N  L  D  K  V  N  S  N  S  L  D  L  P  S  S  S  D  T  T  H  A  S  H  V
```

Tuesday, 18 November 1997 13:58 Page 5
pLM3 (1 > 10847) Site and Sequence

CCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGCACCCATCCTCAATATTAACTCAGCCAGCTTCT 2900
GGTCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGA

insert pLM1

ORF pLM1

P D L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F

CCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGAT 3000
GGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCGGACGTGTCCTCGTACCTCAGGGAGGTCTA

insert pLM1

ORF pLM1

S Q G L E L M S G F S V P K E T R M Y P K L S G L H R S M E S L Q M

GCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTCCCACAGAAGAAGAGACGGAAGAGCTGACT 3100
CGGTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGTCTTCTTCTCTGCCTTCTCGACTGA

insert pLM1

ORF pLM1

P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T

TGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGG 3200
ACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCTCC

insert pLM1

ORF pLM1

W S G S P R A G Q L D S N Q R D R N T L P K K G L R Y Q L Q S Q E

AGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACTTCCCATGTCTCTGAG 3300
TCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTC

insert pLM1

ORF pLM1

E T K E R R H S H T I G G L P E S D D Q S E L P S P P A L P M S L S

TGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACAGCATCCCCACCCACGAGGCGGCCTTCGAG 3400
ACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTC

insert pLM1

ORF pLM1

A K G Q L T N I V S P T A A T T P R I T R S N S I P T H E A A F E

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 6

CTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGGATCCTTCCGAGACCCCACGGACGATGTTC
GACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCCTAGGAAGGCTCTGGGGTGCCTGCTACAAG 3500

insert pLM1

ORF pLM1

L Y S G S Q M G S T L S L A E R P K G M I R S G S F R D P T D D V

ACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAAATCCGGAAGCTTCGTAGGGAACT
TGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGA 3600

insert pLM1

ORF pLM1

H G S V L S L A S S A S S T Y S S A E E R M Q S E Q I R K L R R E L

GGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCC
CCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAAAACTCGTCTCGGACCACTTATACTGTAGG 3700

insert pLM1

ORF pLM1

E S S Q E K V A T L T S Q L S A N A N L V A A F E Q S L V N M T S

CGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCC
GCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTTCTTGAGACTCCGGG 3800

insert pLM1

ORF pLM1

R L R H L A E T A E E K D T E L L D L R E T I D F L K K K N S E A

AGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACTCCTCAGATAGCATCTCAAGCCTCAACAG
TCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCTGTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTC 3900

insert pLM1

ORF pLM1

Q A V I Q G A L N A S E T T P K E L R I K R Q N S S D S I S S L N S

CATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAA
GTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTT 4000

insert pLM1

ORF pLM1

I T S H S S I G S S K D A D A K K K K K K S W V Y E L R S S F N K

Tuesday, 18 November 1997 13:58 Page 7
pLM3 (1 > 10847) Site and Sequence

GCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCCAAACTAC 4100
CGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATG

insert pLM1

ORF pLM1

A F S I K K G P K S A S S Y S D I E E I A T P D S S A P S S P K L

AGCATGGTTCCACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTCACCCAGCCCCCCA 4200
TCGTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGT

insert pLM1

ORF pLM1

Q H G S T E T A S P S I K S S T L S S V G T D V T E G P A H P A P H

CACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGAC 4300
GTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCCTTTACTTCGAATGTCTG

insert pLM1

ORF pLM1

T R L F H A N E E E E P E K K E V S E L R S E L W E K E M K L T D

ATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATG 4400
TAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTAC

insert pLM1

ORF pLM1

I R L E A L N S A H Q L D Q L R E T M H N M Q L E V D L L K A E N

ACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGC 4500
TGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCG

insert pLM1

ORF pLM1

D R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S L G L A

ACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTG 4600
TGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACTGGGAGGCCCAC

insert pLM1

ORF pLM1

L T H S F G P S L A D T D L S P M D G I S T C G P K E E V T L R V

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

GTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGA 4700
CACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCT

insert pLM1

ORF pLM1

V V R M P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D V K

TGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAG 4800
ACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAGGTACCGATGTC

insert pLM1

ORF pLM1

M L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I H G Y S

CATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAG 4900
GTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTC

insert pLM1

ORF pLM1

I S H V K R V L D A E P P E M P P C R R G V N N I S V S L K G L K

GAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCT 5000
CTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGA

insert pLM1

ORF pLM1

E K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L

CGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCAC 5100
GCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTG

insert pLM1

ORF pLM1

S G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T

CTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTG 5200
GAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGAC

insert pLM1

ORF pLM1

F N M H Q Q S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L

Tuesday, 18 November 1997 13:58 Page 9
pLM3 (1 > 10847) Site and Sequence

GTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTA 5300
CACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCAT

insert pLM1

ORF pLM1

V I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y I I G

CCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCT 5400
GGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGA

insert pLM1

ORF pLM1

T T N Q P V K M T P N H G L H L S F R M L T F S N N V E P A N G F L

GGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGG 5500
CCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACC

insert pLM1

ORF pLM1

V R Y L R R K L V E S D S D I N A N K E E L L R V L D W V P K L W

TATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGA 5600
ATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCT

insert pLM1

ORF pLM1

Y H L H T F L E K H S T S D F L I G P C F F L S C P I G I E D F R

CCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGA 5700
GGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCT

insert pLM1

ORF pLM1

T W F I D L W N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D

CCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGC 5800
GGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCG

insert pLM1

ORF pLM1

P V E W V R D T L P W P S A Q Q D Q S K L Y H L P P P T V G P H S

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence
Page 10

ATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTG 5900
TAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGAC

insert pLM1

ORF pLM1

I A S P P E D R T V K D S T P S S L D S D P L M A M L L K L Q E A

CCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGA 6000
GGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCT

insert pLM1

ORF pLM1

A N Y I E S P D R E T I L D P N L Q A T L . G F G N H C H P R T A E

ACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAA 6100
TGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTT

insert pLM1

R W H Q L S . L L L S P L L F Q S T G S P A P G G E Q E G G G D E

AGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTG 6200
TCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGAC

insert pLM1

R G G T G S W C C T F E N F L G R N G G V A F G N L C P L N T F T

GCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGTTCAGAA 6300
CGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTT

insert pLM1

G L L . . L W G K D D S G S F P . L L V S I T N S W A F W G G V Q I

AACATCAAAACACTGCAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGGCTGAACTTGCTCAAAAGAAGCCGAATTCCAGCACACTGGCGGCCGTTA 6400
TTGTAGTTTTGTGACGTCGTCAAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGTTTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAAT

insert pLM1

T S K H C S S S P E F S L D L T R L N L L K R S R I P A H W R P L

CTAGTTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGAC 6500
GATCAAGATCTCCCGGGCAAATTTGGGCGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCACGGAAGGAACTG

L V L E G P F K P A D Q P R L C L L V A S H L L F A P P P C L P .

CCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG 6600
GGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTTACTCCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCCACCCCACCCC

P W K V P L P L S F P N K M R K L H R I V . V G V I L F W G V G V G

CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTA
GTCCTGTCGTTCCCCCTCCTAACCCTTCTGTTATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCCTTTCTTGGTCGACCCCGAGAT 6700
R T A R G R I G K T I A G M L G M R W A L W L L R R K E P A G A L

GGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGC
CCCCCATAGGGGTGCGCGGGACATCGCCGCGTAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCG 6800
G G I P T R P V A A H . A R R V W W L R A A . P L H L P A P . R P

TCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTA
AGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCGTAGGGAAATCCCAAGGCTAAATCACGAAAT 6900
L L S L S S L P F S P R S P A F P V K L . I G A S L . G S D L V L Y

CGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT
GCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCA 7000
G T S T P K N L I R V M V H V V G H R P D R R F F A L . R W S P R

TCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTG
AGAAATTATCACCTGAGAACAAGGTTTGACCTTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACCCCTAAAGCCGGATAAC 7100
S L I V D S C S K L E Q H S T L S R S I L L I Y K G F W G F R P I

GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGC
CAATTTTTTACTCGACTAAATTGTTTTTAAATTGCGCTTAATTAAGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCCGTCCG 7200
G . K M S . F N K N L T R I N S V E C V S V R V W K V P R L P R Q A

AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATT
TCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAA 7300
E V C K A C I S I S Q Q P G V E S P Q A P Q Q A E V C K A C I S I

AGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTAT
TCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATA 7400
S Q Q P . S R P . L R P S R P . L R P V P P I L R P M A D . F F L

TTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCT
AATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTCGAGGGCCCTCGA 7500
F M Q R P R P P L P L S Y S R S S E E A F L E A . A F A K S S R E L

TGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGG
ACATATAGGTAAAAGCCTAGACTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACC 7600
V Y P F S D L I K R Q D E D R F A . L N K M D C T Q V L R P L G W

AGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAA
TCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTT 7700
R G Y S A M T G H N R Q S A A L M P P C S G C Q R R G A R F F L S

GACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTT
CTGGCTGGACAGGCCACGGGACTTACTTGACGTCCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAA 7800
R P T C P V P . M N C R T R Q R G Y R G W P R R A F L A Q L C S T L

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 12

```
GTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGG
CAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACC  7900
 S  L  K  R  E  G  T  G  C  Y  V  A  K  C  R  G  R  I  S  C  H  L  T  L  L  L  P  R  K  Y  P  S  W

CTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGC
GACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCG  8000
 L  M  Q  C  G  G  C  I  R  L  I  R  L  P  A  H  S  T  T  K  R  N  I  A  S  S  E  H  V  L  G  W  K

CGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGAT
GCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGGCTGCCGCTCCTA  8100
 P  V  L  S  I  R  M  I  W  T  K  S  I  R  G  S  R  Q  P  N  C  S  P  G  S  R  R  A  C  P  T  A  R  I

CTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACC
GAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGG  8200
 S  S  .  P  M  A  M  P  A  C  R  I  S  W  W  K  M  A  A  F  L  D  S  S  T  V  A  G  W  V  W  R  T

GCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGA
CGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCT  8300
 A  I  R  T  .  R  W  L  P  V  I  L  L  K  S  L  A  A  N  G  L  T  A  S  S  C  F  T  V  S  P  L  P

TTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAC
AAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTG  8400
 I  R  S  A  S  P  S  I  A  F  L  T  S  S  S  E  R  D  S  G  V  R  N  D  R  P  S  D  A  Q  P  A  I  T

GAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCT
CTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGA  8500
 R  F  R  F  H  R  R  L  L  .  K  V  G  L  R  N  R  F  P  G  R  R  L  D  D  P  P  A  R  G  S  H  A

GGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTG
CCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGAC  8600
 G  V  L  R  P  P  Q  L  V  Y  C  S  L  .  W  L  Q  I  K  Q  .  H  H  K  F  H  K  .  S  !  F  F  T

CATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGT
GTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACA  8700
 A  F  .  L  W  F  V  Q  T  H  Q  C  I  L  S  C  L  Y  T  V  D  L  .  L  E  L  G  V  I  M  V  I  A  V

TTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCAC
AAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACGGATTACTCACTCGATTGAGTG  8800
 S  C  V  K  L  L  S  A  H  N  S  T  Q  H  T  S  R  K  H  K  V  .  S  L  G  C  L  M  S  E  L  T  H

ATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGT
TAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCACGGTCGACGTAATTACTTAGCCGGTTGCGCGCCCCTCTCCGCCAAACGCA  8900
 I  N  C  V  A  L  T  A  R  F  P  V  G  K  P  V  V  P  A  A  L  M  N  R  P  T  R  G  E  R  R  F  A

ATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
TAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATA  9000
 V  W  A  L  F  R  F  L  A  H  .  L  A  A  L  G  R  S  A  A  A  S  G  I  S  S  L  K  G  G  N  T  V  I
```

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence
Page 13

```
CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATC 9100
 H  R  I  R  G  .  R  R  K  E  H  V  S  K  R  P  A  K  G  Q  E  P  .  K  G  R  V  A  G  V  F  P  .

GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGC
CGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCG 9200
 A  P  P  P  .  R  A  S  Q  K  S  T  L  K  S  E  V  A  K  P  D  R  T  I  K  I  P  G  V  S  P  W  K

TCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTA
AGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACAT 9300
L  P  R  A  L  S  C  S  D  P  A  A  Y  R  I  P  V  R  L  S  P  F  G  K  R  G  A  F  S  M  L  T  L  .

GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
CCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGA 9400
 V  S  Q  F  G  V  G  R  S  L  Q  A  G  L  C  A  R  T  P  R  S  A  R  P  L  R  L  I  R  .  L  S  S

TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
ACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAAC 9500
.  V  Q  P  G  K  T  R  L  I  A  T  G  S  S  H  W  .  Q  D  .  Q  S  E  V  C  R  R  C  Y  R  V  L

AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGAT
TTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTA 9600
E  V  V  A  .  L  R  L  H  .  K  D  S  I  V  Y  L  R  S  A  E  A  S  Y  L  R  K  K  S  W  .  L  L  I

CCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTC
GGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAG 9700
 R  Q  T  N  H  R  W  .  R  W  F  F  C  L  Q  A  A  D  Y  A  Q  K  K  R  I  S  R  R  S  F  D  L  F

TACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
ATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTT 9800
Y  G  V  .  R  S  V  E  R  K  L  T  L  R  D  F  G  H  E  I  I  K  K  D  L  H  L  D  P  F  K  L  K

TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
ACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAG 9900
M  K  F  .  I  N  L  K  Y  I  .  V  N  L  V  .  Q  L  P  M  L  N  Q  .  G  T  Y  L  S  D  L  S  I  S

GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACG
CAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGC 10000
 F  I  H  S  C  L  T  P  R  R  V  D  N  Y  D  T  G  G  L  T  I  W  P  Q  C  C  N  D  T  A  R  P  T

CTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT
GAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTA 10100
L  T  G  S  R  F  I  S  N  K  P  A  S  R  K  G  R  A  Q  K  V  S  C  N  F  I  R  L  H  P  V  Y  .

TGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
ACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCAT 10200
L  L  P  G  S  .  S  K  .  F  A  S  .  .  F  A  Q  R  C  C  H  C  Y  R  H  R  G  V  T  L  V  V  W  Y
```

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 1 4

```
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGT
ACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACA  1030
 G  F  I  Q  L  R  F  P  T  I  K  A  S  Y  M  I  P  H  V  V  Q  K  S  G  .  L  L  R  S  S  D  R  C

CAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGT
GTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCA  1040
 Q  K  .  V  G  R  S  V  I  T  H  G  Y  G  S  T  A  .  F  S  Y  C  H  A  I  R  K  M  L  F  C  D  W

GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
CTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTATGCCCTATTATGGCGCGGTGTATCGTCTTGAA  1050
 .  V  L  N  Q  V  I  L  R  I  V  Y  A  A  T  E  L  L  L  P  G  V  N  T  G  .  Y  R  A  T  .  Q  N  F

TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAA
ATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTT  1060
 K  S  A  H  H  W  K  T  F  F  G  A  K  T  L  K  D  L  T  A  V  E  I  Q  F  D  V  T  H  S  C  T  Q

CTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGT
GACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACA  1070
 L  I  F  S  I  F  Y  F  H  Q  R  F  W  V  S  K  N  R  K  A  K  C  R  K  K  G  N  K  G  D  T  E  M

TGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
ACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTG  1080
 L  N  T  H  T  L  P  F  S  I  L  L  K  H  L  S  G  L  L  S  H  E  R  I  H  I  .  M  Y  L  E  K  .  T

AAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC
TTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAG  10847
 N  R  G  S  A  H  I  S  P  K  S  A  T  .  R
```

FIG. 1.

Sequence of pTB72, an expression vector incorporating *C. elegans* UNC-53. The Open reading frame (ORF) of the prolinker + *Ce*-UNC-53 and (upper ORF) *Ce*-UNC-53 alone (lower ORF) are listed under the sequence.

;modified from ptb72.do patent by changing ori on circle;sequence;Created: maandag, 8 juli 1996 09:27::

```
AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG
TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT
CGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTC
TCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTT
GGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTGCAGATATCCATCACACTGGCGG
CCGCCGCCATGACGACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTCGAAG
GGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAAT
GTGATCGTTCCGATCAACGAATTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGG
CCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCTCGAAACTCACCAAAACCGATATCGACA
GCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAAA
AAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGT
GTCGCCACGTCAGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAG
ACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCA
CCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAAT
GTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAAC
CGACCTACCTCCCAACTCCAAAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATC
GGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTGCTTCTGTGAAGACTATTGGAGCAAA
ACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCAT
CTTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATC
GCTGCCCCAGTGAAAAGTGGCCTGAAGCCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTT
GTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAGACTCGAAACGATGCTCAAAGA
GCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCT
AAGCATGCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAA
CGCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTG
TTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATCCACCTCCAGCTGTTCCG
CCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGT
GATATCTGAAAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCAC
CTCTAAAATCAGTTGTTCCACTTAAAATGACTTCAATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGG
AAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACTCCATTGTGGCTCATG
CGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAG
ACATCAGAATCCAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAG
AATACGATGACATGACTCGTCGAGCACAGAACGGCTATCCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCT
GGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGGCAACAGTCGC
CTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTC
GGTCCTCCACATCAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGA
ACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGGACAACATTCGCTAAGATCCCCGGGATACTCATCCTA
TTCTCCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTATGCACTCACAGACTAGTCGACGACCTTCTTC
ACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATCCACCGA
GCACAGAATGGCGGCTCTCTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCC
TACTCGGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGA
TGAAAAATCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGCTAGCACGACAGCATATG
GATCTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGTC
GACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAAACTCACT
CAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACGATTCAGGCAGGACATTGCTCATTTGAGGGA
TATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCTTCGTCAACCATCTCT
GGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGA
GCTCGTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAAC
TACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAACTCTTGACAACATTGATGTGATTGAGTTGAAG
CAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTTGA
TGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCA
ACGGTCCAGCCACTCGTGCTTCTTCCCGCGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCA
GCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAAC
GTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTC
AACCAGTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGAT
GTGGAGCATCAACTTGGAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATT
GGAGACTCCACAACCATGATAACCAGCCATCCAACTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCATG
CACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACT
CGTCAAGTCAATTTTGACAGAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAACTGGCGA
AGACCCTGGCTGCTTATGTATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAA
CAATAAAGAAGAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGCAAAGAATCATGCATCGTAAT
TCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGTCCCACTTCAAAACAACGAA
GGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAAATGTCA
GTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGACGACGGGCGGTAGAGGATGAGTATCGTCT
AACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCTCTTCAGGCCGTCAATAATTTT
ATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACTGTCCTCTAACTGTCGATGGA
```

FIG. 1 CONTINUED.

TCCCGTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAA
AAAACCTTCGGTCGCTGCACTTCCTTCGAGGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGG
TGAAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAAGACCTCGTCCCGTCACCTGCCAACTCATCCCGAC
AACACTTCAATCCCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAG
AAGACTCTAATCTTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCC
CCCTTTTCCCCCCAATTTCCCAGAACCTCCTGTTCCCTTTGTTCCTAGTCCTCCCGGGTGCCGACGCCGAAGCG
ATTTAAAAACCTTTTTCTTTCCGAAACATTTCCCATTGCTCATTAATAGTCAAATTGAATAAACAGTGTATGTACTT
AAAAAAAAAAAAAAAAAAACTCGAGGGGGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCTGATCAG
CCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGT
GCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGT
GGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGG
CGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGC
TCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCC
CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGT
GGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTT
CCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTA
TTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGT
GGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGT
GGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC
GCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTT
TTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGA
GGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGG
ATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGC
TATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCG
GTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGC
TGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATT
GGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGAT
GCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGC
GAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC
CAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATG
CCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGC
GGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGC
TTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTT
CTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCA
CCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGG
GGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAG
CATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTT
ATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATT
GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG
AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTA
ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCG
CTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAAT
CAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGA
AACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCT
GCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGG
TATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCT
GCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCT
TGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA
AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAA
AGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCC
CAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGA
GTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTC
GTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA
AAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATG
GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAG
TCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA
TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCT
CCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGC
TTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATT
GCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACAT
TGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTA
CATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTAT
GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTT
GGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTA

FIG. 2.

Met Thr Thr Ser Asn Val Glu Leu Ile Pro Ile
Tyr Thr Asp Trp 15

Ala Asn Arg His Leu Ser Lys Gly Ser Leu Ser
Lys Ser Ile Arg 30

Asp Ile Ser Asn Asp Phe Arg Asp Tyr Arg Leu
Val Ser Gln Leu 45

Ile Asn Val Ile Val Pro Ile Asn Glu Phe Ser
Pro Ala Phe Thr 60

Lys Arg Leu Ala Lys Ile Thr Ser Asn Leu Asp
Gly Leu Glu Thr 75

Cys Leu Asp Tyr Leu Lys Asn Leu Gly Leu Asp
Cys Ser Lys Leu 90

Thr Lys Thr Asp Ile Asp Ser Gly Asn Leu Gly
Ala Val Leu Gln 105

Leu Leu Phe Leu Leu Ser Thr Tyr Lys Gln Lys
Leu Arg Gln Leu 120

Lys Lys Asp Gln Lys Lys Leu Glu Gln Leu Pro
Thr Ser Ile Met 135

Pro Pro Ala Val Ser Lys Leu Pro Ser Pro Arg
Val Ala Thr Ser 150

Ala Thr Ala Ser Ala Thr Asn Pro Asn Ser Asn
Phe Pro Gln Met 165

Ser Thr Ser Arg Leu Gln Thr Pro Gln Ser Arg
Ile Ser Lys Ile 180

Asp Ser Ser Lys Ile Gly Ile Lys Pro Lys Thr
Ser Gly Leu Lys 195

Pro Pro Ser Ser Ser Thr Thr Ser Ser Asn Asn
Thr Asn Ser Phe 210

Arg Pro Ser Ser Arg Ser Ser Gly Asn Asn Asn
Val Gly Ser Thr 225

Ile Ser Thr Ser Ala Lys Ser Leu Glu Ser Ser
Ser Thr Tyr Ser 240

FIG. 2 CONTINUED.

Ser Ile Ser Asn Leu Asn Arg Pro Thr Ser Gln
Leu Gln Lys Pro 255

Ser Arg Pro Gln Thr Gln Leu Val Arg Val Ala
Thr Thr Thr Lys 270

Ile Gly Ser Ser Lys Leu Ala Ala Pro Lys Ala
Val Ser Thr Pro 285

Lys Leu Ala Ser Val Lys Thr Ile Gly Ala Lys
Gln Glu Pro Asp 300

Asn Ser Gly Gly Gly Gly Gly Gly Met Leu Lys
Leu Lys Leu Phe 315

Ser Ser Lys Asn Pro Ser Ser Ser Ser Asn Ser
Pro Gln Pro Thr 330

Arg Lys Ala Ala Ala Val Pro Gln Gln Gln Thr
Leu Ser Lys Ile 345

Ala Ala Pro Val Lys Ser Gly Leu Lys Pro Pro
Thr Ser Lys Leu 360

Gly Ser Ala Thr Ser Met Ser Lys Leu Cys Thr
Pro Lys Val Ser 375

Tyr Arg Lys Thr Asp Ala Pro Ile Ile Ser Gln
Gln Asp Ser Lys 390

Arg Cys Ser Lys Ser Ser Glu Glu Glu Ser Gly
Tyr Ala Gly Phe 405

Asn Ser Thr Ser Pro Thr Ser Ser Ser Thr Glu
Gly Ser Leu Ser 420

Met His Ser Thr Ser Ser Lys Ser Ser Thr Ser
Asp Glu Lys Ser 435

Pro Ser Ser Asp Asp Leu Thr Leu Asn Ala Ser
Ile Val Thr Ala 450

Ile Arg Gln Pro Ile Ala Ala Thr Pro Val Ser
Pro Asn Ile Ile 465

Asn Lys Pro Val Glu Glu Lys Pro Thr Leu Ala
Val Lys Gly Val 480

Lys Ser Thr Ala Lys Lys Asp Pro Pro Pro Ala
Val Pro Pro Arg 495

Asp Thr Gln Pro Thr Ile Gly Val Val Ser Pro
Ile Met Ala His 510

FIG. 2 CONTINUED.

Lys Lys Leu Thr Asn Asp Pro Val Ile Ser Glu
Lys Pro Glu Pro 525

Glu Lys Leu Gln Ser Met Ser Ile Asp Thr Thr
Asp Val Pro Pro 540

Leu Pro Pro Leu Lys Ser Val Val Pro Leu Lys
Met Thr Ser Ile 555

Arg Gln Pro Pro Thr Tyr Asp Val Leu Leu Lys
Gln Gly Lys Ile 570

Thr Ser Pro Val Lys Ser Phe Gly Tyr Glu Gln
Ser Ser Ala Ser 585

Glu Asp Ser Ile Val Ala His Ala Ser Ala Gln
Val Thr Pro Pro 600

Thr Lys Thr Ser Gly Asn His Ser Leu Glu Arg
Arg Met Gly Lys 615

Asn Lys Thr Ser Glu Ser Ser Gly Tyr Thr Ser
Asp Ala Gly Val 630

Ala Met Cys Ala Lys Met Arg Glu Lys Leu Lys
Glu Tyr Asp Asp 645

Met Thr Arg Arg Ala Gln Asn Gly Tyr Pro Asp
Asn Phe Glu Asp 660

Ser Ser Ser Leu Ser Ser Gly Ile Ser Asp Asn
Asn Glu Leu Asp 675

Asp Ile Ser Thr Asp Asp Leu Ser Gly Val Asp
Met Ala Thr Val 690

Ala Ser Lys His Ser Asp Tyr Ser His Phe Val
Arg His Pro Thr 705

Ser Ser Ser Ser Lys Pro Arg Val Pro Ser Arg
Ser Ser Thr Ser 720

Val Asp Ser Arg Ser Arg Ala Glu Gln Glu Asn
Val Tyr Lys Leu 735

Leu Ser Gln Cys Arg Thr Ser Gln Arg Gly Ala
Ala Ala Thr Ser 750

Thr Phe Gly Gln His Ser Leu Arg Ser Pro Gly
Tyr Ser Ser Tyr 765

Ser Pro His Leu Ser Val Ser Ala Asp Lys Asp
Thr Met Ser Met 780

FIG. 2 CONTINUED.

His Ser Gln Thr Ser Arg Arg Pro Ser Ser Gln
Lys Pro Ser Tyr 795

Ser Gly Gln Phe His Ser Leu Asp Arg Lys Cys
His Leu Gln Glu 810

Phe Thr Ser Thr Glu His Arg Met Ala Ala Leu
Leu Ser Pro Arg 825

Arg Val Pro Asn Ser Met Ser Lys Tyr Asp Ser
Ser Gly Ser Tyr 840

Ser Ala Arg Ser Arg Gly Gly Ser Ser Thr Gly
Ile Tyr Gly Glu 855

Thr Phe Gln Leu His Arg Leu Ser Asp Glu Lys
Ser Pro Ala His 870

Ser Ala Lys Ser Glu Met Gly Ser Gln Leu Ser
Leu Ala Ser Thr 885

Thr Ala Tyr Gly Ser Leu Asn Glu Lys Tyr Glu
His Ala Ile Arg 900

Asp Met Ala Arg Asp Leu Glu Cys Tyr Lys Asn
Thr Val Asp Ser 915

Leu Thr Lys Lys Gln Glu Asn Tyr Gly Ala Leu
Phe Asp Leu Phe 930

Glu Gln Lys Leu Arg Lys Leu Thr Gln His Ile
Asp Arg Ser Asn 945

Leu Lys Pro Glu Glu Ala Ile Arg Phe Arg Gln
Asp Ile Ala His 960

Leu Arg Asp Ile Ser Asn His Leu Ala Ser Asn
Ser Ala His Ala 975

Asn Glu Gly Ala Gly Glu Leu Leu Arg Gln Pro
Ser Leu Glu Ser 990

Val Ala Ser His Arg Ser Ser Met Ser Ser Ser
Ser Lys Ser Ser 1005

Lys Gln Glu Lys Ile Ser Leu Ser Ser Phe Gly
Lys Asn Lys Lys 1020

Ser Trp Ile Arg Ser Ser Leu Ser Lys Phe Thr
Lys Lys Lys Asn 1035

Lys Asn Tyr Asp Glu Ala His Met Pro Ser Ile
Ser Gly Ser Gln 1050

FIG. 2 CONTINUED.

Gly Thr Leu Asp Asn Ile Asp Val Ile Glu Leu
Lys Gln Glu Leu 1065

Lys Glu Arg Asp Ser Ala Leu Tyr Glu Val Arg
Leu Asp Asn Leu 1080

Asp Arg Ala Arg Glu Val Asp Val Leu Arg Glu
Thr Val Asn Lys 1095

Leu Lys Thr Glu Asn Lys Gln Leu Lys Lys Glu
Val Asp Lys Leu 1110

Thr Asn Gly Pro Ala Thr Arg Ala Ser Ser Arg
Ala Ser Ile Pro 1125

Val Ile Tyr Asp Asp Glu His Val Tyr Asp Ala
Ala Cys Ser Ser 1140

Thr Ser Ala Ser Gln Ser Ser Lys Arg Ser Ser
Gly Cys Asn Ser 1155

Ile Lys Val Thr Val Asn Val Asp Ile Ala Gly
Glu Ile Ser Ser 1170

Ile Val Asn Pro Asp Lys Glu Ile Ile Val Gly
Tyr Leu Ala Met 1185

Ser Thr Ser Gln Ser Cys Trp Lys Asp Ile Asp
Val Ser Ile Leu 1200

Gly Leu Phe Glu Val Tyr Leu Ser Arg Ile Asp
Val Glu His Gln 1215

Leu Gly Ile Asp Ala Arg Asp Ser Ile Leu Gly
Tyr Gln Ile Gly 1230

Glu Leu Arg Arg Val Ile Gly Asp Ser Thr Thr
Met Ile Thr Ser 1245

His Pro Thr Asp Ile Leu Thr Ser Ser Thr Thr
Ile Arg Met Phe 1260

Met His Gly Ala Ala Gln Ser Arg Val Asp Ser
Leu Val Leu Asp 1275

Met Leu Leu Pro Lys Gln Met Ile Leu Gln Leu
Val Lys Ser Ile 1290

Leu Thr Glu Arg Arg Leu Val Leu Ala Gly Ala
Thr Gly Ile Gly 1305

Lys Ser Lys Leu Ala Lys Thr Leu Ala Ala Tyr
Val Ser Ile Arg 1320

FIG. 2 CONTINUED.

Thr Asn Gln Ser Glu Asp Ser Ile Val Asn Ile
Ser Ile Pro Glu 1335

Asn Asn Lys Glu Glu Leu Leu Gln Val Glu Arg
Arg Leu Glu Lys 1350

Ile Leu Arg Ser Lys Glu Ser Cys Ile Val Ile
Leu Asp Asn Ile 1365

Pro Lys Asn Arg Ile Ala Phe Val Val Ser Val
Phe Ala Asn Val 1380

Pro Leu Gln Asn Asn Glu Gly Pro Phe Val Val
Cys Thr Val Asn 1395

Arg Tyr Gln Ile Pro Glu Leu Gln Ile His His
Asn Phe Lys Met 1410

Ser Val Met Ser Asn Arg Leu Glu Gly Phe Ile
Leu Arg Tyr Leu 1425

Arg Arg Arg Ala Val Glu Asp Glu Tyr Arg Leu
Thr Val Gln Met 1440

Pro Ser Glu Leu Phe Lys Ile Ile Asp Phe Phe
Pro Ile Ala Leu 1455

Gln Ala Val Asn Asn Phe Ile Glu Lys Thr Asn
Ser Val Asp Val 1470

Thr Val Gly Pro Arg Ala Cys Leu Asn Cys Pro
Leu Thr Val Asp 1485

Gly Ser Arg Glu Trp Phe Ile Arg Leu Trp Asn
Glu Asn Phe Ile 1500

Pro Tyr Leu Glu Arg Val Ala Arg Asp Gly Lys
Lys Thr Phe Gly 1515

Arg Cys Thr Ser Phe Glu Asp Pro Thr Asp Ile
Val Ser Lys Lys 1530

Trp Pro Trp Phe Asp Gly Glu Asn Pro Glu Asn
Val Leu Lys Arg 1545

Leu Gln Leu Gln Asp Leu Val Pro Ser Pro Ala
Asn Ser Ser Arg 1560

Gln His Phe Asn Pro Leu Glu Ser Leu Ile Gln
Leu His Ala Thr 1575

Lys His Gln Thr Ile Asp Asn Ile

FIG. 3. : **tblastn search of the EST division of Genbank with the ORF of the longest known *Ce*-UNC-53 cDNA, tb3-M5, reveals two EST's with homology to a predicted coiled-coil region in *Ce*-UNC-53.**

```
TBLASTN 1.4.9MP [26-March-1996] [Build 14:27:13 Apr 1
1996]

Reference:  Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W.
Myers, and David J. Lipman (1990).  Basic local alignment search tool.  J. Mol.
Biol. 215:403-10.

Query=  tb3_M5 ORF
        (1583 letters)
Database:  Non-redundant Database of GenBank EST Division
           647,253 sequences; 234,216,808 total letters.
                                                                       Smallest
                                                                         Sum
                                                      Reading  High  Probability
Sequences producing High-scoring Segment Pairs:         Frame Score  P(N)      N
dbj|D35780|CELK025D6F C.elegans cDNA clone yk25d6 : 5'... +2   358  7.9e-54    3
dbj|D33048|CELK025D6R C.elegans cDNA clone yk25d6 : 3'... -1   177  8.6e-16    1
gb|H09036|H09036      yl96c11.r1 Homo sapiens cDNA clo... +1   115  1.1e-05    1
gb|AA049124|AA049124  mj46f04.r1 Soares mouse embryo N... +3   106  8.6e-05    1
gb|R91475|R91475      yq08c11.r1 Homo sapiens cDNA clo... +2    59  0.21       2
gb|T23446|T23446      seq2955 Homo sapiens cDNA clone ... -1    61  0.99       2
gb|R86390|R86390      SW3ICA339SK Brugia malayi infect... +2    74  0.996      1
gb|T44781|T44781      8044 Arabidopsis thaliana cDNA c... +1    71  0.9992     1
gb|T75582|T75582      yd63f11.r1 Homo sapiens cDNA clo... +2    64  0.99992    2

gb|H09036|H09036 yl96c11.r1 Homo sapiens cDNA clone 46037 5'.
            Length = 489

  Plus Strand HSPs:
 Score = 115 (52.1 bits), Expect = 1.1e-05, P = 1.1e-05
 Identities = 22/70 (31%), Positives = 45/70 (64%), Frame = +1

Query:  1059 IELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKLTNGPATRA 1118
             ++L+ EL+++   L ++RL+ L  A ++D LRE +N++++E ++LK   D+L +
Sbjct:     7 MQLRNELRDKEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLKVXNDRLKSESQGSG 186

Query:  1119 SSRASIPVIY 1128
              SR S P ++
Sbjct:   187 CSRGSFPSVH 216


gb|AA049124|AA049124 mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus
            musculus cDNA clone 479167 5'
            Length = 337
  Plus Strand HSPs:
 Score = 106 (48.0 bits), Expect = 8.6e-05, P = 8.6e-05
 Identities = 23/58 (39%), Positives = 38/58 (65%), Frame = +3

Query:  1057 DVIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKLTNGP 1114
             +V EL+ EL E++  L ++RL+ L+ A ++D LRET++ ++ E   LK E D+L   P
Sbjct:    99 EVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLKVAP 272
```

FIG. 4

A Search of the Genbank databases with part of the nucleotide binding domain of *Ce*-UNC-53 does not identify statistically significant proteins except for the C. elegans cosmid containing *Ce-unc-53*.

```
TBLASTN 1.4.8MP [20-June-1995] [Build 18:00:05 Aug 29 1995]
Query= section5         (240 letters)
>lcl|section5
ILTERRLVLAGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPENNKEELLQVERRLE
KILRSKESCIVILDNIPKNRIAFVVSVFANVPLQNNEGPFVVCTVNRYQIPELQIHHNFK
MSVMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFFPIALQAVNNFIEKTNSVD
VTVGPRACLNCPLTVDGSREWFIRLWNENFIPYLERVARDGKKNLRSLHFLRGSHRHRLX

Database:  Non-redundant PDB+GBupdate+GenBank+EMBLupdate+EMBL
           520,383 sequences; 367,017,413 total letters.

                                                                 Smallest
                                                                   Sum
                                                   Reading  High  Probability
Sequences producing High-scoring Segment Pairs:      Frame Score  P(N)       N
emb|Z47810|CEF45E10  Caenorhabditis elegans cosmid F45... -2  1131  5.1e-158   2
gb|R41071|R41071     Hk575-f Homo sapiens cDNA clone k... +2    53  0.33       2
gb|T44781|T44781     8044 Arabidopsis thaliana cDNA cl... +1    74  0.35       1
emb|Z48334|CEF10B5   Caenorhabditis elegans cosmid F10... +3    71  0.83       3
gb|M81884|EPFCPCG    Epifagus virginiana chloroplast c... +1    49  0.91       4
gb|L09547|PEAPCLP    Pisum sativum (clone pCLp) nuclea... +1    71  0.99       1
gb|M32604|TOMCD4B    Tomato ATP-dependent protease (CD... +1    71  0.99       1
emb|X69188|APTUBGA   A.pyhllitidis mRNA for gamma-tubulin +2    56  0.992      3
gb|T44782|T44782     8045 Arabidopsis thaliana cDNA cl... +1    68  0.9995     1
gb|M17087|HUMRASK12  Human c-ras-Ki-2 activated oncoge... +2    58  0.9998     1
emb|X57702|GGNATRIUP G.gallus RNA for precursor of nat... +3    56  0.9999     2
gb|K01520|HUMRASKB1  Human lung adenocarcinoma (PR371)... +2    57  0.99995    1

>gb|R41071|R41071 Hk575-f Homo sapiens cDNA clone k575-f.
            Length = 310
  Plus Strand HSPs:

 Score = 53 (24.5 bits), Expect = 0.40, Sum P(2) = 0.33
 Identities = 9/15 (60%), Positives = 13/15 (86%), Frame = +2

Query:    130 GFILRYLRRRAVEDE 144
              GF++RYLRR+ VE +
Sbjct:    26 GFLVRYLRRKLVESD 70

 Score = 47 (21.7 bits), Expect = 0.40, Sum P(2) = 0.33
 Identities = 9/26 (34%), Positives = 17/26 (65%), Frame = +3

Query:    170 NNFIEKTNSVDVTVGPRACLNCPLTV 195
              + F+EK +++D  +GP   L+ PL +
Sbjct:    147 HTFLEKHSTLDFLIGPCFFLSGPLAL 224
```

FIG. 5.

Three frame translation of EST gb:R41071.

Regions of homology region with *Ce*-Unc-53 in two different frames are underlined. The spacing between the blocks of homology is of similar size as that in *Ce*-UNC-53. Subsequent re-cloning and re-sequencing of this region in man identified multiple sequencing errors gb:R41071, and identified an ORF which is more homologous to and co-linear with *Ce*-UNC-53 (see alignment in fig 12).

```
CTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGT
          10        20        30        40        50        60
 L  Q  Q  R  G  A  S  Q  W  L  P  G  S  L  P  E  E  E  A  G
  S  N  N  V  E  P  A  N  G  F  L  V  R  Y  L  R  R  K  L  V
   P  T  T  W  S  Q  P  M  A  S  W  F  V  T  *  G  G  S  W  *

AGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGGTGCTCGACTTGG
          70        80        90       100       110       120
 R  V  R  Q  R  H  Q  C  Q  Q  G  R  A  A  S  G  C  S  T  W
  E  S  D  S  D  I  N  A  N  K  E  E  L  L  R  G  A  R  L  G
   S  Q  T  A  T  S  M  P  T  R  K  S  C  F  G  V  L  D  L  G

GTACCCAAGCCTGTGGTATCATCTTCCACACCTTCCTTGAGAAGCACAGCACCTTAGACT
         130       140       150       160       170       180
 V  P  K  P  V  V  S  S  S  T  P  S  L  R  S  T  A  P  *  T
  Y  P  S  L  W  Y  H  L  P  H  L  P  *  E  A  Q  H  L  R  L
   T  Q  A  C  G  I  I  F  H  T  F  L  E  K  H  S  T  L  D  F

TTCTCATCGGCCCTTGCTTCTTTCTGTCGGGTCCATTGGCATTGAGGCTTCCGGACCTTG
         190       200       210       220       230       240
 F  S  S  A  L  A  S  F  C  R  V  H  W  H  *  G  F  R  T  L
  S  H  R  P  L  L  L  S  V  G  S  I  G  I  E  A  S  G  P  C
   L  I  G  P  C  F  F  L  S  G  P  L  A  L  R  L  P  D  L  V

TTTATTGACCTGTGGACAACTCTATCATTTCCTATCTACAGGAGGAGCCAAGGATTGGAT
         250       260       270       280       290       300
 F  I  D  L  W  T  T  L  S  F  P  I  Y  R  R  S  Q  G  L  D
  L  L  T  C  G  Q  L  Y  H  F  L  S  T  G  G  A  K  D  W  I
   Y  *  P  V  D  N  S  I  I  S  Y  L  Q  E  E  P  R  I  G  *

AAAGGTCCAT
         310
 K  G  P
  K  V  H
   R  S
```

FIG. 6: **blastn search of the EST division of Genbank with Hu-unc-53/1 cDNA 3b.**

```
BLASTN 1.4.9MP [26-March-1996] [Build 14:27:07 Apr  1 1996]
Query=  Hu-unc-53/1 cDNA 3b
        (3256 letters)
Database:  Non-redundant Database of GenBank EST Division
           647,253 sequences; 234,216,808 total letters.
```

Sequences producing High-scoring Segment Pairs:		High Score	Smallest Sum Probability P(N)	N	LOCUS assignment
gb|N36659|N36659	yx91b09.r1 Homo sapiens cDNA clone 2...	1668	2.1e-130	1	hu-UNC-53/1
gb|AA043997|AA043997	zk58a01.r1 Soares pregnant uterus Nb...	1316	8.3e-129	3	hu-UNC-53/1
gb|AA049124|AA049124	mj46f04.r1 Soares mouse embryo NbME1...	1324	9.1e-102	1	mu-UNC-53/1
gb|T05560|T05560	EST03449 Homo sapiens cDNA clone HFB...	892	5.1e-84	3	hu-UNC-53/1
gb|N24681|N24681	yx91b09.s1 Homo sapiens cDNA clone 2...	782	9.9e-75	2	hu-UNC-53/1
gb|R41071|R41071	Hk575-f Homo sapiens cDNA clone k575-f.	535	1.5e-72	4	hu-UNC-53/1
gb|N89104|N89104	K7846F Fetal heart, Lambda ZAP Expre...	451	7.3e-57	2	hu-UNC-53/1
gb|R41073|R41073	Hk144-f Homo sapiens cDNA clone k144-f.	555	1.5e-36	1	hu-UNC-53/1
gb|R15492|R15492	HH434-F Homo sapiens cDNA clone H434-F.	416	2.3e-29	2	hu-UNC-53/1
gb|H09036|H09036	y196c11.r1 Homo sapiens cDNA clone 4...	438	9.4e-26	1	hu-UNC-53/2
gb|W91567|W91567	MTA.C36.093.A MTA adult mouse thymus...	317	1.9e-17	2	mu-UNC-53/?
gb|W74400|W74400	zd62e10.r1 Soares fetal heart NbHH19...	243	2.2e-09	1	hu-UNC-53/1
gb|AA003314|AA003314	mg56h10.r1 Soares mouse embryo NbME1...	141	0.54	1	

FIG. 7.

TBLASTN search of the Genbank sequence database with the 961 aminoacid ORF of cDNA 3b of hu-UNC-53/1. hu-UNC-53/1forms a unique pair with *Ce*-UNC-53 (cosmid F45E10) compared to the rest of the database.

```
TBLASTN 1.4.9MP (26-March-1996) [Build 14:27:13 Apr  1 1996]

Query=  tmpseq_1
        (961 letters)

Database:  Non-redundant GenBank+EMBL+DDBJ+PDB sequences
           261,674 sequences; 371,416,172 total letters.
```

Sequences producing High-scoring Segment Pairs:		Reading Frame	High Score	Smallest Sum Probability P(N)	N
emb|Z47810|CEF45E10	Caenorhabditis elegans cosmid F45E10	-2	158	2.3e-32	7
gb|M97501|HUMCLIP	Human cytoplasmic linker protein-...	+3	83	0.47	1
emb|X64838|HSRESTIN	H.sapiens mRNA for restin	+1	83	0.47	1
gb|M58752|ECOMCRBC	E.coli mcrB and mcrC genes, compl...	+3	82	0.56	1
emb|Z11582|SCNUF1G	S.cerevisiae nufl gene	+1	82	0.61	2
emb|X73297|SCSETRP4	S.cerevisiae spacer element	+1	82	0.74	2
emb|X54002|XLKINESIN	X.laevis mRNA for kinesine	+2	63	0.85	5
gb|U42409|DDU42409	Dictyostelium discoideum myosin h...	+3	66	0.92	2
gb|U10399|YSCH8082	Saccharomyces cerevisiae chromoso...	+2	77	0.93	2
gb|U20810|ATU20810	Arabidopsis thaliana cytoskeleton...	+1	77	0.95	2
gb|L07879|LEIKINLIKE	Leishmania chagasi kinesin-like p...	+2	78	0.95	1
gb|L03188|YSCINTANA	Saccharomyces cerevisiae integrin...	+2	65	0.997	2
gb|U28372|YSCD9476	Saccharomyces cerevisiae chromoso...	+3	82	0.9991	2
gb|M94362|HUMLAMBBA	Human lamin B2 (LAMB2) mRNA, part...	+1	75	0.9996	1
gb|M58337|VACHAGMA	Vaccinia virus hemagglutinin gene.	+1	74	0.99995	1

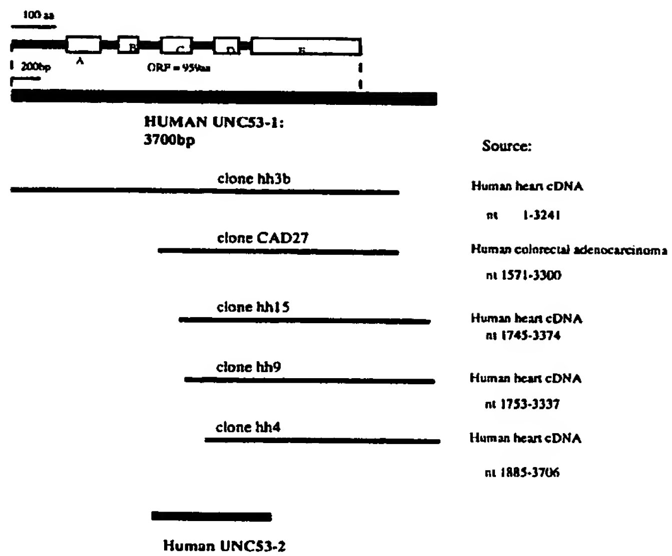

100 aa
B
C
D
E
200bp
A
ORF = 959aa
HUMAN UNC53-1:
3700bp
Source:
clone hh3b
Human heart cDNA
nt 1-3241
clone CAD27
Human colorectal adenocarcinoma
nt 1571-3300
clone hh15
Human heart cDNA
nt 1745-3374
clone hh9
Human heart cDNA
nt 1753-3337
clone hh4
Human heart cDNA
nt 1885-3706
Human UNC53-2

Figure 8

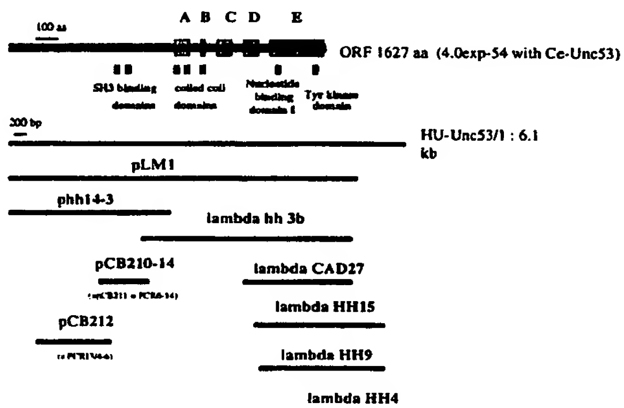

A B C D E
100 aa
ORF 1627 aa (4.0exp-54 with Ce-Unc53)
SH3 binding domains
coiled coil domains
Nucleotide binding domain I
Tyr kinase domain
200 bp
HU-Unc53/1 : 6.1 kb
pLM1
phh14-3
lambda hh 3b
pCB210-14
lambda CAD27
lambda HH15
pCB212
lambda HH9
lambda HH4

Figure 8a

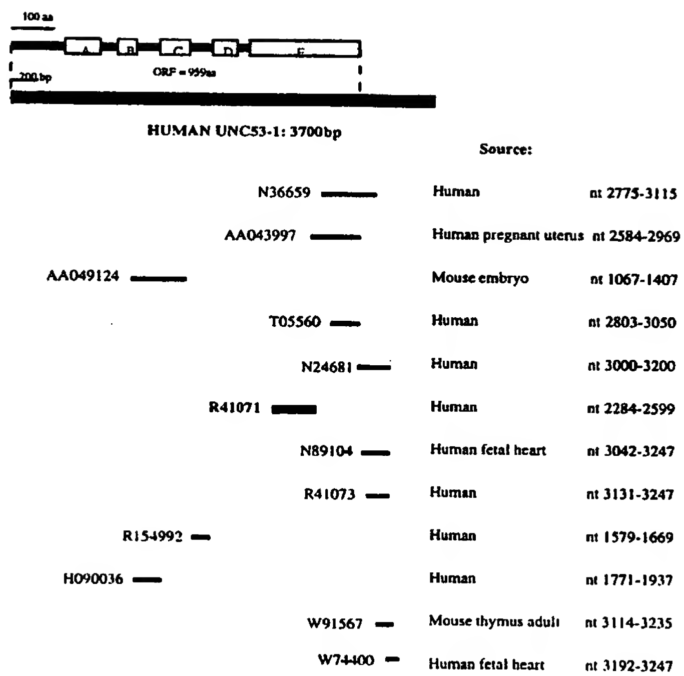

100 aa
A
B
C
D
E
200 bp
ORF = 959aa
HUMAN UNC53-1: 3700bp
Source:
N36659 Human nt 2775-3115
AA043997 Human pregnant uterus nt 2584-2969
AA049124 Mouse embryo nt 1067-1407
T05560 Human nt 2803-3050
N24681 Human nt 3000-3200
R41071 Human nt 2284-2599
N89104 Human fetal heart nt 3042-3247
R41073 Human nt 3131-3247
R154992 Human nt 1579-1669
H090036 Human nt 1771-1937
W91567 Mouse thymus adult nt 3114-3235
W74400 Human fetal heart nt 3192-3247

Figure 8b

FIG. 9a

```
GAATTCCGGGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGG 100
 N  S  G  E  E  L  T  V  S  G  S  P  R  A  G  Q  L  D  S  N  Q  R  D  R  N  T  L  P  K  K  G  L  R
 Not part of hum
TACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCC 200
 Y  Q  L  Q  S  Q  E  E  T  K  E  R  R  H  S  H  T  I  G  G  L  P  E  S  D  D  Q  S  E  L  P  S  P
CTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACAGCATCCC 300
 P  A  L  P  M  S  L  S  A  K  G  Q  L  T  N  I  V  S  P  T  A  A  T  T  P  R  I  T  R  S  N  S  I  P
CACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGGATCCTTC 400
 T  H  E  A  A  F  E  L  Y  S  G  S  Q  M  G  S  T  L  S  L  A  E  R  P  K  G  M  I  R  S  G  S  F
CGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAAA 500
 R  D  P  T  D  D  V  H  G  S  V  L  S  L  A  S  S  A  S  S  T  Y  S  S  A  E  E  R  M  Q  S  E  Q
                                                homology block A
TCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCA 600
 I  R  K  L  R  R  E  L  E  S  S  Q  E  K  V  A  T  L  T  S  Q  L  S  A  N  A  N  L  V  A  A  F  E  Q
                                                homology block A
GAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGACTTTCTG 700
 S  L  V  N  M  T  S  R  L  R  H  L  A  E  T  A  E  E  K  D  T  E  L  L  D  L  R  E  T  I  D  F  L
                                                homology block A
AAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACTCCTCAG 800
 K  K  K  N  S  E  A  Q  A  V  I  Q  G  A  L  N  A  S  E  T  T  P  K  E  L  R  I  K  R  Q  N  S  S
                            homology block A
ATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGGTCTATGA 900
 D  S  I  S  S  L  N  S  I  T  S  H  S  S  I  G  S  S  K  D  A  D  A  K  K  K  K  K  K  S  W  V  Y  E
                                                                             homology block B
GCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCA 1000
 L  R  S  S  F  N  K  A  F  S  I  K  K  G  P  K  S  A  S  S  Y  S  D  I  E  E  I  A  T  P  D  S  S
 homology block B
GCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTCACCGAGG 1100
 A  P  S  S  P  K  L  Q  H  G  S  T  E  T  A  S  P  S  I  K  S  S  T  L  S  S  V  G  T  D  V  T  E
GCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGA 1200
 G  P  A  H  P  A  P  H  T  R  L  F  H  A  N  E  E  E  E  P  E  K  K  E  V  S  E  L  R  S  E  L  W  E
                                                                  homology block C
GAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTG 1300
 K  E  M  K  L  T  D  I  R  L  E  A  L  N  S  A  H  Q  L  D  Q  L  R  E  T  M  H  N  M  Q  L  E  V
                                                homology block C
GACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCC 1400
 D  L  L  K  A  E  N  D  R  L  K  V  A  P  G  P  S  S  G  S  T  P  G  Q  V  P  G  S  S  A  L  S  S
                            homology block C
                                                      weak homology in hum1 vs hum2
CACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAA 1500
 P  R  R  S  L  G  L  A  L  T  H  S  F  G  P  S  L  A  D  T  D  L  S  P  M  D  G  I  S  T  C  G  P  K
                                     weak homology in hum1 vs hum2
GGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAGCAAGGTC 1600
 E  E  V  T  L  R  V  V  V  R  M  P  P  Q  H  I  I  K  G  D  L  K  Q  Q  E  F  F  L  G  C  S  K  V
                                                                             homology block D
                            weak homology in hum1 vs hum2
AGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCA 1700
 S  G  K  V  D  W  K  M  L  D  E  A  V  F  Q  V  F  K  D  Y  I  S  K  M  D  P  A  S  T  L  G  L  S
                                                homology block D
CTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATC 1800
 T  E  S  I  H  G  Y  S  I  S  H  V  K  R  V  L  D  A  E  P  P  E  M  P  P  C  R  R  G  V  N  N  I  S
                   homology block D
```

FIG. 9a CONTINUED.

AGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTG 1900
V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L L L
homology block E - pred nucleotide 8D

AAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGG 2000
K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G R E
homology block E - pred nucleotide 8D

TCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAAC 2100
V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D R E T
homology block E - pred nucleotide 8D

AGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCAT 2200
G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C K Y H
homology block E - pred nucleotide 8D

AAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACG 2300
K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S N N
homology block E - pred nucleotide 8D

TGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCT 2400
V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L R V L
homology block E - pred nucleotide 8D

CGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCC 2500
D W V P K L W Y H L H T F L E K H S T S D F L I G P C F F L S C P
homology block E - pred nucleotide 8D

ATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATG 2600
I G I E D F R T W F I D L W N N S I I P Y L Q E G A K D G I K V H
homology block E - pred nucleotide 8D

GACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCC 2700
G Q K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H L P P
homology block E - pred nucleotide 8D

ACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATG 2800
P T V G P H S I A S P P E D R T V K D S T P S S L D S D P L M A M
homology block E - pred. nucleotide 8D

CTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATC 2900
L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L
homology block E - pred nucleotide 8D

ACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGA 3000
3' untranslated trailer

ACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTG 3100
3' untranslated trailer

TGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGG 3200
3' untranslated trailer

CTTTCTGGGGAGGGGTTCAGAAAACATCCAAACACTGCAGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGGAGATCTG 3300
3' untranslated trailer

GGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCCATCACCACTGCCAACAACTCCTCCCCCAGAGATCTGGCTGGAGCCCAGAAA 3400
3' untranslated trailer

AAGAAGCATGTGGTTTAAAAAATGTTTGAATCAATCTGTAAAAGGTAAAAATGAAAAAACAAAAACAAGCAAACAAACAAAAAACAATGGAAAAGATGAA 3500
3' untranslated trailer

GCTGGAGAGAGAGGAACCAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACCTGAGAAG 3600
3' untranslated trailer

TCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAGCTTGGAC 3700
3' untranslated trailer

TTAACC 3706

Fig. 9b

Tuesday, 18 November 1997 10:33 Page 1

. lu-Unc53/1 seq (1 > 6013) Site and Sequence

Enzymes : 60 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

fig 9 14 pag

GATATCTGCAGAATTCGGCTTCTTTGAGCAAGTTCAGCCTGGTTAAGTCCAAGCTGAATTCCGGGGAAAGCCGAGCCGGATCCCTCGAGGACCCTATGCG 100
CTATAGACGTCTTAAGCCGAAGAAACTCGTTCAAGTCGGACCAATTCAGGTTCGACTTAAGGCCCCTTTCGGCTCGGCCTAGGGAGCTCCTGGGATACGC

pCR2.1 linker — lambda gt10 primer EcoRI — suspect sequence linker? — pHH14-3

I S A E F G F F E Q V Q P G . V Q A E F R G K P S R I P R G P Y A

GAGGTCAAGCCGCTCAGCAAGGCGCCTGAAGCGGCCGTGAGCGAAGATGGCAAATCGGACGACGAGCTGCTCTCCAGCAAGGCCAAGGCGCAAAAGAGCT 200
CTCCAGTTCGGCGAGTCGTTCCGCGGACTTCGCCGGCACTCGCTTCTACCGTTTAGCCTGCTGCTCGACGAGAGGTCGTTCCGGTTCCGCGTTTTCTCGA

pHH14-3

E V K P L S K A P E A A V S E D G K S D D E L L S S K A K A Q K S

CTGGGCCTGTCCCCTCTGCCAAGGGCCAGGAGGAGCGCGCCTTCCTCAAGGTGGACCCCGAGCTGGTGGTGACCGTGCTGGGAGACCTGGAGCAGCTGCT 300
GACCCGGACAGGGGAGACGGTTCCCGGTCCTCCTCGCGCGGAAGGAGTTCCACCTGGGGCTCGACCACCACTGGCACGACCCTCTGGACCTCGTCGACGA

pHH14-3

S G P V P S A K G Q E E R A F L K V D P E L V V T V L G D L E Q L L

CTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGA 400
GAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGTCTTGGACCTTCTCTGGTACAGGTCGGACGCT

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S L R

GGGTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCCAACCGCTCGTCCCCTC 500
CCCAGGGTCCACTGAGTGTCGAGGGACCTCTACTGGACGATGCTGTCGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAG

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

G S Q V T H S S L E M T C Y D S D D A N P R S V S S L S N R S S P

TGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTACAT 600
ACAGTACCGCGATACCGGTCAGGTCAGGCGCCGACGTCCGACCACTGCGCGGGAGACACCCACCCTCGACGGCGAGCCTCCCCTGCGGGCGGACCATGTA

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

L S W R Y G Q S S P R L Q A G D A P S V G G S C R S E G T P A W Y M

Tuesday, 18 November 1997 10:33 Fig 9b

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence Page 7

```
GCACGGCGAACGGGCCCACTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTGGACTCG
CGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGC 700
pHH14-3
ORF (1-579bp) = pLM7 ORF
full available ORF HU-Unc53/1 = pLM1 OR
 H  G  E  R  A  H  Y  S  H  T  M  P  M  R  S  P  S  K  L  S  H  I  S  R  L  E  L  V  E  S  L  D  S

GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCA
CTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTACTTTCGT 800
pHH14-3
ORF (1-579bp) = pLM7 ORF
full available ORF HU-Unc53/1 = pLM1 OR
 D  E  V  D  L  K  S  G  Y  M  S  D  S  D  L  M  G  K  T  M  T  E  D  D  D  I  T  T  G  W  D  E  S

GCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC
CGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACG 900
pHH14-3
pCB212
ORF (1-579bp) = pLM7 ORF
full available ORF HU-Unc53/1 = pLM1 OR
 S  S  I  S  S  G  L  S  D  A  S  D  N  L  S  S  E  E  F  N  A  S  S  S  L  N  S  L  P  S  T  P  T  A

TTCTCGCAGGAACTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCC
AAGAGCGTCCTTGAGTTGTTATCACGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCCTCTTTCGG 1000
pHH14-3
pCB212
full available ORF HU-Unc53/1 = pLM1 OR
 S  R  R  N  S  T  I  V  L  R  T  D  S  E  K  R  S  L  A  E  S  G  L  S  W  F  S  E  S  E  E  K  A

CCTAAAAAACTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGAGAGCTGTGATGATTCATCCAAGG
GGATTTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCCTGAAGATTCACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCC 1100
pHH14-3
pCB212
full available ORF HU-Unc53/1 = pLM1 OR
 P  K  K  L  E  Y  D  S  G  S  L  K  M  E  P  G  T  S  K  W  R  R  E  R  P  E  S  C  D  D  S  S  I
```

Tuesday, 18 November 1997 10:33 Fig 9b Page

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCCACCTGTGGCTGTAACTTCCCCCATCACTCACACAGC 1200
CACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCG
                                        pHH14-3
                                        pCB212
                     full available ORF HU-Unc53/1 = pLM1 OR
G  G  E  L  K  K  P  I  S  L  G  H  P  G  S  L  K  K  G  K  T  P  P  V  A  V  T  S  P  I  T  H  T  A

CCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGAT 1300
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGGAGACTA
                                        pHH14-3
                                        pCB212
                     full available ORF HU-Unc53/1 = pLM1 OR
 Q  S  A  L  K  V  A  G  K  P  E  G  K  A  T  D  K  G  K  L  A  V  K  N  T  G  L  Q  R  S  S  S  D

GCTGGTCGGGACCGCCTGAGTGATGCTAAGAAGCCCCCCTCGGGCATTGCTCGCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCA 1400
CGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT
                                        pHH14-3
                                        pCB212
                     full available ORF HU-Unc53/1 = pLM1 OR
 A  G  R  D  R  L  S  D  A  K  K  P  P  S  G  I  A  R  P  S  T  S  G  S  F  G  Y  K  K  P  P  P  A

CAGGCACAGCCACTGTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC 1500
GTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCTG
                                        pHH14-3
                                        pCB212
                     full available ORF HU-Unc53/1 = pLM1 OR
T  G  T  A  T  V  M  Q  T  G  G  S  A  T  L  S  K  I  Q  K  S  S  G  I  P  V  K  P  V  N  G  R  K  T

TAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGT 1600
ATCGAATCTACAAAGGTTGTCACGTCTCGGTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCA
                                        pHH14-3
                                        pCB212
                     full available ORF HU-Unc53/1 = pLM1 OR
 S  L  D  V  S  N  S  A  E  P  G  F  L  A  P  G  A  R  S  N  I  Q  Y  R  S  L  P  R  P  A  K  S  S
```

Tuesday, 18 November 1997 10:33

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

TCTATGAGCGTGACCGGCGGGCGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCTTCCA
AGATACTCGCACTGGCCGCCCGCCCCACCTGGAGCGGGACACTCGTCGTCGTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGGT 1700

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HU53rv4

S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T P S

GACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGC
CTGACTTCCTCGGATGGTTCCATCGGTCACCCGCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCCTCTTCCGGTTTCGGTTCCGTCACCG 1800

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

R L K E P T K V A S G R T T P A P V N Q T D R E K E K A K A K A V A

CTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAAGAACCAAGCAAGCCACCCCACAGCCACCAAGCTGGCAGAGCTG
GAACCTGAGTCTGTTGTAGAGGAACTTCTCATAACCGAGGGGTCTTTCATGAGGGTTCTTGGTTCGTTCGGTGGGGTGTCGGTGGTTCGACCGTCTCGAC 1900

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A E L

CCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCAT
GGTGGTTGGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTTGTCAGACCTAGATGGTAGTA 2000

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L P S

Tuesday, 18 November 1997 10:33 Fig 9b Page 5

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
CCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGCACC  2100
GGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGG
                              pHH14-3
                              pCB210-14
                              full available ORF HU-Unc53/1 = pLM1 OR
 S  S  D  T  T  H  A  S  K  V  P  D  L  H  A  T  S  S  A  S  G  G  P  L  P  S  C  F  T  P  S  P  A  P

CATCCTCAATATTAACTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTCTCAGGC  2200
GTAGGAGTTATAATTGAGTCGGTCGAAGAGGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCG
                              pHH14-3
                              pCB210-14
                              full available ORF HU-Unc53/1 = pLM1 OR
 I  L  N  I  N  S  A  S  F  S  Q  G  L  E  L  M  S  G  F  S  V  P  K  E  T  R  M  Y  P  K  L  S  G

CTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACCTGCTCCCCCTGCTGCTC  2300
GACGTGTCCTCGTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAG
                              pHH14-3
                              pCB210-14
                              full available ORF HU-Unc53/1 = pLM1 OR
 L  H  R  S  M  E  S  L  Q  M  P  M  S  L  P  S  A  F  P  S  S  T  P  V  P  T  P  P  A  P  P  A  A

CCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGG  2400
GGTGTCTTCTTCTCTGCCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCC
                              pHH14-3
                                     pHH3b
                              pCB210-14
                              full available ORF HU-Unc53/1 = pLM1 OR
                    rev primer HU53rv3            rev primer HU53rv2
                                                     peptide B72628H
 P  T  E  E  E  T  E  E  L  T  V  S  G  S  P  R  A  G  Q  L  D  S  N  Q  R  D  R  N  T  L  P  K  K  G
```

Tuesday, 18 November 1997 10:33 Fig 9b

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence Page 6

GCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT
CGAGTCCATGGTCGAAGTCAGGGTCCTCCTCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGA 2500

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L P

TCTCCCCCTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCACGCCAAGAATCACCCGCTCCAACA
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGT 2600

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S N

GCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGG
CGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC 2700

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R S G

ATCCTTCCGAGACCCCACGGACGATGTTCACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCT
TAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGA 2800

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M Q S

GAGCAAATCCGGAAGCTTCGTAGGGAACTGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTT
CTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA 2900

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

E Q I R K L R R E L E S S Q E K V A T L T S Q L S A N A N L V A A

Fig 9b

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 7

TTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGA
AACTCGTCTCGGACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCT 3000

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAAC
GAAAGACTTCTTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCTGTTTTG 3100

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q N

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTTCTTCTTTTTTTTCTCAACCC 3200

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S W

TCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGA
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCT 3300

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D

CTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTTCTACAGAGACTGCTTCACCCTCCATCAAGTCCTCCACCTTGTCCTCCGTGGGCACTGATGTC
GAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGATGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAG 3400

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q H G S T E T A S P S I K S S T L S S V G T D V

Tuesday, 18 November 1997 10:33 Fig 9b Page 8
file Hu-Unc53/1 seq (1 > 6013) Site and Sequence

ACCG GCCCTGCTCACCCAGCCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGC 3500
TGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCCTCCATAGCCTCGACGCGAGACTCG

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTT 3600
ATACCCTCTTCCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAA

U2 ORF = pCB251 ORF

pHH3b

peptide B72627H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTA 3700
CCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAAT

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTG 3800
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACAC

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S T C

Tuesday, 18 November 1997 10:34

fig 'u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

GTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCCTGGGCTGTAG

CAGGTTTCCTCCTTCACTGGGAGGCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATC 3900

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGA

GTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCT 4000

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATA

GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT 4100

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V N

ACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCT

TGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGA 4200

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L

Tuesday, 18 November 1997 10:34 Fig 9b Page 10
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

CCTCCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGC 4300
GGAGGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCG

U2 ORF = pCB251 ORF
pHH3b
U4 ORF = pCB201 ORF
full available ORF HU-Unc53/1 = pLM1 OR
U3 ORF = pLM5 ORF
pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACC 4400
GCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF
pHH3b
U4 ORF = pCB201 ORF
full available ORF HU-Unc53/1 = pLM1 OR
U3 ORF = pLM5 ORF
pHH15

R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

GGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAA 4500
CCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTT

U2 ORF = pCB251 ORF
pHH3b
U4 ORF = pCB201 ORF
full available ORF HU-Unc53/1 = pLM1 OR
U3 ORF = pLM5 ORF
pHH15

R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C H

Tuesday, 18 November 1997 10:34
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence Fig 9b Page 1 1

GTATCAAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCC
CATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGG 4600

U2 ORF = pCB251 ORF
pHH3b
U4 ORF = pCB201 ORF
full available ORF HU-Unc53/1 = pLM1 OR
U3 ORF = pLM5 ORF
pHH15
peptide B72626H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTC
TTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAG 4700

U2 ORF = pCB251 ORF
pHH3b
U4 ORF = pCB201 ORF
full available ORF HU-Unc53/1 = pLM1 OR
U3 ORF = pLM5 ORF
pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTC
CCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAG 4800

U2 ORF = pCB251 ORF
pHH3b
U4 ORF = pCB201 ORF
full available ORF HU-Unc53/1 = pLM1 OR
U3 ORF = pLM5 ORF
pHH15

R V L D W V P K L W Y H L H T F L E K H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34 Fig 9b Page 12
fig .Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAG
CACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTC 4900

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L W N N S I I P Y L Q E G A K D G I K

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACC
CAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGG 5000

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V H G Q K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H

TGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGAT
ACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTA 5100

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S D P L I

Tuesday, 18 November 1997 10:34 Fig 9b Page 13

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

```
GGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTC
CCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAG  5200
```

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72625H

```
A  M  L  L  K  L  Q  E  A  A  N  Y  I  E  S  P  D  R  E  T  I  L  D  P  N  L  Q  A  T  L  .  G  F
```

```
GGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAG
CCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTC  5300
```

pHH3b

pHH15

```
G  N  H  C  H  P  R  T  A  E  R  W  H  Q  L  S  .  L  L  L  S  P  L  L  F  Q  S  T  G  S  P  A  P
```

```
GAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGG
CTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACC  5400
```

pHH3b

pHH15

```
G  G  E  Q  E  G  G  G  D  E  R  G  G  T  G  S  V  C  C  T  F  E  N  F  L  G  R  N  G  G  V  A  F  G
```

```
GAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAAC
CTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTG  5500
```

pHH3b

pHH15

```
N  L  C  P  L  N  T  F  T  G  L  L  .  .  L  V  G  K  D  D  S  G  S  F  P  .  L  L  V  S  I  T  N
```

```
TCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGG
AGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAGTTTTGTGACGTCGTCAAGGATTTACTAAGAGTGTTCGTTGGGACTCTCTCTGTCAGAACACTCCC  5600
```

pHH3b

pHH15

```
S  W  A  F  V  G  G  V  Q  K  T  S  K  H  C  S  S  S  .  M  I  L  T  S  N  P  E  R  D  S  L  V  R
```

Tuesday, 18 November 1997 10:34
fig. Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 4

```
AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCCATCACCACTGCCAACAACTCCTCCCCCAGAGATCTGGCTGGAGC
TCTAGACCCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGTCTGGGAAGGGTTAAGGTAGTGGTGACGGTTGTTGAGGAGGGGGTCTCTAGACCGACCTCG  5700
                          pHH15
 E  I  W  G  R  Q  E  A  P  Q  I  F  S  Q  T  L  P  N  S  I  T  T  A  N  N  S  S  P  R  D  L  A  G  A

CCAGAAAAAGAAGCATGTGGTTTAAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAAACAAAAACAAGCAAACAAACAAAAAACAATGGAAA
GGTCTTTTTCTTCGTACACCAAATTTTTTACAAATTTAGTTAGACATTTTCCATTTTTACTTTTTTGTTTTTGTTCGTTTGTTTGTTTTTTGTTACCTTT  5800
  Q  K  K  K  H  V  V  .  K  M  F  K  S  I  C  K  R  .  K  .  K  N  K  N  K  Q  T  N  K  K  Q  V  I

AGATGAAGCTGGAGAGAGAGGAACCAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACC
TCTACTTCGACCTCTCTCTCCTTGGTCAACGGTTCCATCTCTCGACGGGCGAGGACGGGAGACCTACTGTATCCCCTGTAGTTGTTCTGCCGACGGTTGG  5900
 R  .  S  V  R  E  R  N  Q  L  P  R  .  R  A  A  R  S  C  P  L  D  D  I  G  D  I  N  K  T  A  A  N

TGAGAAGTCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAG
ACTCTTCAGTGGTTTGGTGTTTTTATTGGAATGTCGGAAGTCCCTTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTGTTACGTGGCCTTAAGTC  6000
                                                                                          linker?
 L  R  S  H  Q  T  T  K  I  T  L  Q  P  S  G  K  D  Y  Q  L  C  L  S  T  L  .  F  N  N  A  P  E  F  S

CTTGGACTTAACC  6013
GAACCTGAATTGG
  linker?
 L  D  L  T
```

FIG. 10.

GGCACGAGGCATCCTCTGTGGGCACCGAGGTCACCGAGACCCCTGCTCATTCAGTCCCCCACACTAGACT 70

linker ? open reading frame

R H E A S S V G T E V T E T P A H S V P H T R L

GTTCCAAGCCAATGAAGAGGAGGAGCCAGAGAAGAAGGAGGTATCAGAACTGCGCTCTGAACTATGGGAA 140

open reading frame

F Q A N E E E E P E K K E V S E L R S E L V E

AAAGAGATGAAGCTCACGGATATCCGGTTGGAGGCCCTCAACTCTGCCCACCAGCTGGACCAGCTTCGGG 210

open reading frame

K E M K L T D I R L E A L N S A H Q L D Q L R

AGACCATGCACAATATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGGCTGAAGGTTGCCCC 280

open reading frame

E T M H N M Q L E V D L L K A E N D R L K V A P

CGGCCCCTCCTCAGGCTGCACTCCAGGGCAGGTCCCTGGGTCATCGGCTCTGTCGTCCCCTCGACGTTCC 350

open reading frame

G P S S G C T P G Q V P G S S A L S S P R R S

CTGGGCCTTGCACTCAGCCATCCTTTCAGTCCTAGTCTCACAGACACAGACCTCTCACCCATGGATGGCA 420

open reading frame

L G L A L S H P F S P S L T D T D L S P M D G

TCAGCACCTGTGGTTCAAAGGAAGAGGTGACCCTGCGGGTGGTGGTCCGGATGCCGCCCCAGCACATCAT 490

open reading frame

I S T C G S K E E V T L R V V V R M P P Q H Q I

CAAAGGGGACTTAAAGCAGCAGGAGTTCTTCCTGGGTTGCAGCAAGGTCAGTGGCAAAGTTGACTGGAAG 560

open reading frame

K G D L K Q Q E F F L G C S K V S G K V D W K

ATGCTGGATGAAGCCGTTTTCCAAGTGTTCAAGGACTACATTTCTAAAATGGACCCAGCCTCAACCCTGG 630

open reading frame

M L D E A V F Q V F K D Y I S K M D P A S T L

GACTGAGCACTGAGTCCATACATGGCTATAGCCTCAGCCACGTGAAACGAGTGCTGGATGCTGAGCCCCC 700

open reading frame

G L S T E S I H G Y S L S H V K R V L D A E P P

FIG. 10 CONTINUED.

AGAGATGCCTCCTTGCCGCCGAGGTGTCAATAACATATCAGTCGCTCTCAAAGGTCTGAAAGAGAAGTGT 770
open reading frame
E M P P C R R G V N N I S V A L K G L K E K C

GTCGACAGCCTGGTGTTCGAGACGCTTATCCCCAAGCCCATGATGCAGCACTACATCAGCCTCCTGCTCA 840
open reading frame
V D S L V F E T L I P K P M M Q H Y I S L L L

AGCACCGGCGCCTGGTGCTCTCCGGCCCCAGTGGCACCGGCAAGACCTACTTGACCAATCGGCTAGCCGA 910
open reading frame
K H R R L V L S G P S G T G K T Y L T N R L A E

GTACCTGGTGGAGCGCTCCGGCCGCGAGGTCACGGATGGCATCGTCAGCACTTTCAACATGCACCAGCAG 980
open reading frame
Y L V E R S G R E V T D G I V S T F N M H Q Q

TCTTGCAAGGATCTGCAACTGTACCTCTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGGATAGGGG 1050
open reading frame
S C K D L Q L Y L S N L A N Q I D R E T G I G

ATGTGCCCTTGGTGATCCTCCTGGATGATCTGAGTGAAGCAGGCTCCATCAGTGAGCTGGTCAATGGGGC 1120
open reading frame
D V P L V I L L D D L S E A G S I S E L V N G A

CCTCACCTGCAAGTATCACAAATGTCCCTACATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCC 1190
open reading frame
L T C K Y H K C P Y I I G T T N Q P V K M T P

AACCATGGCTTGCACTTGAGCTTCAGGATGCTGACCTTCTCGAACAATGTGGAACCAGCCAATGGCTTTC 1260
open reading frame
N H G L H L S F R M L T F S N N V E P A N G F

TGGTCCGTTACCTGCGGAGGAAGTTGGTAGAGTCAGACAGTGACGTCAATGCTAACAAGGAAGAGCTGCT 1330
open reading frame
L V R Y L R R K L V E S D S D V N A N K E E L L

TCGGGTGCTGGACTGGGTGCCCAAGCTGTGGTATCACCTCCACACCTTCCTGGAGAAGCACAGCACCTCG 1400
open reading frame
R V L D W V P K L W Y H L H T F L E K H S T S

FIG. 10 CONTINUED.

GACTTCCTCATTGGCCCTTGCTTCTTCCTGTCCTGTCCCATTGGCATCGAGGACTTCCGGACCTGGTTCA 1470
open reading frame

D F L I G P C F F L S C P I G I E D F R T W F .

TTGACCTGTGGAACAATTCCATCATCCCCTATCTACAGGAAGGAGCCAAGGATGGGATCAAGGTTCATGG 1540
open reading frame

I D L W N N S I I P Y L Q E G A K D G I K V H G

ACAGAAAGCTGCTTGGGAAGACCCGGTGGAATGGGTCCGAGACACTCTTCCCTGGCCGTCGGCCCAACAA 1610
open reading frame

Q K A A W E D P V E W V R D T L P W P S A Q Q

GACCAATCAAAGCTCTACCACCTGCCCCCGCCTTCTGTGGGCCCCCACAGCACTGCCTCACCCCCGGAGG 1680
open reading frame

D Q S K L Y H L P P P S V G P H S T A S P P E

ACAGGACAGTCAAAGACAGCACTCCAAACTCCCTCGACTCAGATCCCCTGATGGCCATGCTACTGAAACT 1750
open reading frame

D R T V K D S T P N S L D S D P L M A M L L K L

CCAAGAAGCTGCCAACTACATTGAGTCACCAGATCGAGAGACTATCCTGGACCCCAACCTCCAGGCGACA 1820
open reading frame

Q E A A N Y I E S P D R E T I L D P N L Q A T

CTCTGAGGGCCCGGCAGTCACTGTCACCCTGGAGGGCAGAAGGCTGGCTTCAGCATCATTAGCTCTCCTC 1890
3' untranslated

L . G P G S H C H P G G Q K A G F S I I S S P

TGCCCTCTTCCTTCATAGCTCTGGCTCACCAGCCTCGCCAAGAGAACAGGAGGGAAGAAGAGGGCAGGAG 1960
3' untranslated

L P S S F I A L A H Q P R Q E N R R E E E G R R

GAGGGATGGGTTCTCGGTGCTGAACCTTTGAGAACTTCCTACTAGGAATTGGAGGGGGTGGAGTTTGAGA 2030
3' untranslated

R D G F S V L N L . E L P T R N W P G W S L R

ACTCCGTGCCCCTTAACTACATTTGCTGGCCTCCTCTTACGACTTAGGAGAAAAGATGATTCTGGTCTTT 2100
3' untranslated

T P C P L T T F A G L L L R L R R K D D S G L

TCTTCAAGTTTTGTTTCACCTACAAACTCTTGGGCTTTCTGGGGAGGGATTCGGAAGATATAAACAGACA 2170
3' untranslated

F F K F C F T Y K L L G F L G R D S E D I N P Q

FIG. 10 CONTINUED.

AACAAAAACAAACAAACCAACTACAGCAGTTCCAAGCTCGTTCTCACAAACACCTCTGAGACAGTCACAT 2240
3' untranslated
T K T N K P T T A V P S S F S Q T P L R Q S H

GTGGGCAAATCTAAGGGAGGCAGGAAGCTCTACAGACTTTCTTGCAAACCCTTCCCAGTTCTGTCGACAC 2310
3' untranslated
V G K S K G G R K L Y R L S C K P F P V L S T

TGCCAACAACCTCCCCGCCAGAGACCTGGCCAGAGCCAAGAAAAGAGAAGCATGTGGTTTAACAGAAAAA 2380
3' untranslated
L P T T S P P E T W P E P R K E K H V V . Q K N

CAAAACAAAACAAAACAAAAAATATATGTGTAAATCAACCTGTAGAAGGTAAAAACGGCAATGGAAAAGA 2450
3' untranslated
K T K Q N K K Y M C K S T C R R . K R Q W K R

TGAAGCTGGAAGGAGGGGCCCAGTTGCCAAGATGGAACGAGAGCTGCCAGATCTTGCCTTCTGGATGACA 2520
3' untranslated
. S V K E G P S C Q D G T R A A R S C L L D D

AGAGGGGACATTGCAAGATGGCTGCCAGTCTAAAACGTCACCAGACCACAAGAGTAACATCACAGCCTTC 2590
3' untranslated
K R G H C K M A A S L K R H Q T T R V T S Q P S

GAAGAAAGGCCACAAGCTGTCTTTCTGCCCTCTAACTGAACATGCATGAAAAGTCAATAAACCCTACTTT 2660
3' untranslated
K K G H K L S F C P L T E H A . K V N K P Y F

TTAATTTTTAAAAAAAAAAAAAAAAAAAAAAAAAATTTCCGCGGCCGC 2709
polyA tail + linker
L I F K K K K K K K K K F P R P

FIG. 11a.

AAGCTTGGCACGAGGCCTCGTGCCAAGCTGAGACCGTCATGCAGCTCCGAAATGAGTTAAGAGACAAGGA 70

LINKER ? open reading frame

A V H E A S C Q A E T V M Q L R N E L R D K E

GATGAAGCTGACAGATATCCGCTTAGAAGCTCTCAGTTCTGCCCACCAGCTGGACCAGCTCCGGGAGGCC 140

open reading frame

M K L T D I R L E A L S S A H Q L D Q L R E A

ATGAACAGGATGCAGAGTGAAATAGAGAAGCTGAAAGCTGAGAATGATCGGCTGAAGTCAGAGTCTCAAG 210

open reading frame

M N R M Q S E I E K L K A E N D R L K S E S Q

GCAGTGGCTGCAGCCGGGCTCCTTCCCAAGTGTCCATCTCTGCCTCCCCGAGGCAGTCCATGGGCCTCTC 280

open reading frame

G S G C S R A P S Q V S I S A S P R Q S M G L S

CCAGCACAGCTTGAACCTCACTGAGTCAACCAGCCTGGACATGTTGCTGGATGACACTGGTGAATGCTCG 350

open reading frame

Q H S L N L T E S T S L D M L L D D T G E C S

GCTCGGAAGGAAGGAGGCAGGCATGTTAAGATAGTTGTCAGCTTTCAGGAGGAAATGAAGTGGAAGGAGG 420

open reading frame

A R K E G G R H V K I V V S F Q E E M K V K E

ATTCCAGACCACACCTCTTTCTTATTGGCTGCATTGGAGTTAGTGGCAAGACGAAGTGGATGTGCTCGA 490

open reading frame

D S R P H L F L I G C I G V S G K T K V D V L D

TGGGGTGGTTAGACGGCTGTTCAAAGAATACATCATTCATGTCGACCCAGTGAGTCAGCTAGGGCTGAAT 560

open reading frame

G V V R R L F K E Y I I H V D P V S Q L G L N

TCAGACAGCGTTCTTGGCTACAGCATTGGAGAAATCAAGCGCAGCAACACTTCCGAAACACCGGAGCTGC 630

open reading frame

S D S V L G Y S I G E I K R S N T S E T P E L

TTCCTTGTGGCTATCTGGTTGGAGAGAACACGACCATCTCAGTGACTGTGAAAGGGCTCGCAGAAAACAG 700

open reading frame

L P C G Y L V G E N T T I S V T V K G L A E N S

CCTGGACTCACTGGTGTTTGAGTCCTTGATTCCCAAGCCCATCCTGCAGCGCTACGTCTCCCTCCTGATA 770

open reading frame

L D S L V F E S L I P K P I L Q R Y V S L L I

GAGCACCGTCGGATCATTCTCTCTGGCCCCAGCGCACTGGGAAAACCTACCTGGCCAACCGGCTGTCTG 840

open reading frame

E H R R I I L S G P S G T G K T Y L A N R L S

FIG. 11a CONTINUED.

AGTATATAGTGCTTCGAGAGGGACGGGAGTTGACAGACGGGGTTATCGCCACCTTTAACGTGGACCATAA 910
open reading frame
E Y I V L R E G R E L T D G V I A T F N V D H K

GTCCAGCAAGGAATTGCGCCAGTACCTGTCCAACCTTGCTGACCAGTGCAACAGTGAGAACAATGCTGTG 980
open reading frame
S S K E L R Q Y L S N L A D Q C N S E N N A V

GACATGCCCCTCGTCATCATCCTGGACAACCTACACCACGTGAGCTCTCTGGGCGAGATCTTCAATGGGC 1050
open reading frame
D M P L V I I L D N L H H V S S L G E I F N G

TGCTCAACTGCAAGTACCACAAATGCCCTTACATAATTGGCACAATGAACCAGGCTACCTATCTCCCCTT 1120
open reading frame
STRAT [illegible]
L L N C K Y H K C P Y I I G T M N Q A T Y L P F

TTATACTAATAATCTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTCTGCGGCCGC 1190
open reading frame
LINKER-vector
Y T N N L I K K K K K K K K K K K K K F C G R

FIG. 11b.

	10 20 30 40 50
1 1...50 UNC53_huma 5 1...48	AAGCTTGGCA CGAGGCCTCG TGCCAAGCTG AGACCGTCAT GCAGCTCCGA A W H E A S C Q A E T V M Q L R LINKER > ORF (START)
	60 70 80 90 100
1 51...100 UNC53_huma 5 49...98	AATGAGTTAA GAGACAAGGA GATGAAGCTG ACAGATATCC GCTTAGAAGC N E L R D K E M K L T D I R L E A
	110 120 130 140 150
1 101...150 UNC53_huma 5 99...148	TCTCAGTTCT GCCCACCAGC TGGACCAGCT CCGGGAGGCC ATGAACAGGA L S S A H Q L D Q L R E A M N R
	160 170 180 190 200
1 151...200 UNC53_huma 5 149...198	TGCAGAGTGA AATAGAGAAG CTGAAAGCTG AGAATGATCG GCTGAAGTCA M Q S E I E K L K A E N D R L K S
	210 220 230 240 250
1 201...250 UNC53_huma 5 199...248	GAGTCTCAAG GCAGTGGCTG CAGCCGGGCT CCTTCCCAAG TGTCCATCTC E S Q G S G C S R A P S Q V S I S
	260 270 280 290 300
1 251...300 UNC53_huma 5 249...298	TGCCTCCCCG AGGCAGTCCA TGGGCCTCTC CCAGCACAGC TTGAACCTCA A S P R Q S M G L S Q H S L N L
	310 320 330 340 350
1 301...350 UNC53_huma 5 299...348	CTGAGTCAAC CAGCCTGGAC ATGTTGCTGG ATGACACTGG TGAATGCTCG T E S T S L D M L L D D T G E C S
	360 370 380 390 400
1 351...400 UNC53_huma 5 349...398	GCTCGGAAGG AAGGAGGCAG GCATGTTAAG ATAGTTGTCA GCTTTCAGGA A R K E G G R H V K I V V S F Q E
	410 420 430 440 450
1 401...450 UNC53_huma 5 399...448	GGAAATGAAG TGGAAGGAGG ATTCCAGACC ACACCTCTTT CTTATTGGCT E M K W K E D S R P H L F L I G
	460 470 480 490 500
1 451...500 UNC53_huma 5 449...498	GCATTGGAGT TAGTGGCAAG ACGAAGTGGG ATGTGCTCGA TGGGGTGGTT C I G V S G K T K W D V L D G V V
	510 520 530 540 550
1 501...550 UNC53_huma 5 499...548	AGACGGCTGT TCAAAGAATA CATCATTCAT GTCGACCCAG TGAGTCAGCT R R L F K E Y I I H V D P V S Q L
	560 570 580 590 600
1 551...600 UNC53_huma 5 549...598	AGGGCTGAAT TCAGACAGCG TTCTTGGCTA CAGCATTGGA GAAATCAAGC G L N S D S V L G Y S I G E I K
	610 620 630 640 650
1 601...650 UNC53_huma 5 599...646	GCAGCAACAC TTCCGAAACA CCGGAGCTGC TTCCTTGTGG CTATCTGGTT R S N T S E T P E L L P C G Y L V
	660 670 680 690 700
1 651...700 UNC53_huma 5 649...696	GGAGAGAACA CGACCATCTC AGTGACTGTG AAAGGGCTCG CAGAAAACAG G E N T T I S V T V K G L A E N S

FIG. 11b CONTINUED.

	710 720 730 740 750
1 701..750 UNC53_huma	CCTGGACTCA CTGGTGTTTG AGTCCTTGAT TCCCAAGCCC ATCCTGCAGC
5 699..748	L D S L V F E S L I P K P I L Q
	760 770 780 790 800
1 751..800 UNC53_huma	GCTACGTCTC CCTCCTGATA GAGCACCGTC GGATCATTCT CTCTGGCCCC
5 749..798	R Y V S L L I E H R R I I L S G P
	810 820 830 840 850
1 801..850 UNC53_huma	AGCGGCACTG GGAAAACCTA CCTGGCCAAC CGGCTGTCTG AGTATATAGT
5 799..848	S G T G K T Y L A N R L S E Y I V
	860 870 880 890 900
1 851..900 UNC53_huma	GCTGCGAGAG GGACGGGAGT TGACAGACGG GGTTATCGCC ACCTTTAACG
5 849..898	L R E G R E L T D G V I A T F N
	910 920 930 940 950
1 901..950 UNC53_huma	TGGACCATAA GTCCAGCAAG GAATTGCGCC AGTACCTGTC CAACCTTGCT
5 899..948	V D H K S S K E L R Q Y L S N L A
	960 970 980 990 1000
1 951..1000 UNC53_hum	GACCAGTGCA ACAGTGAGAA CAATGCTGTG GACATGCCCC TCGTCATCAT
5 949..998	D Q C N S E N N A V D M P L V I I
	1010 1020 1030 1040 1050
1 1001..1050 UNC53_hu	CCTGGACAAC CTACACCACG TGAGCTCTCT GGGCGAGATC TTCAATGGGC
5 999..1048	L D N L H H V S S L G E I F N G
	1060 PRIMER 1070 1080 1090 1100
1 1051..1100 UNC53_hu	TGCTCAACTG CAAGTACCAC AAATGCCCTT ACATAATTGG CACAATGAAC
5 1049..1098	L L N C K Y H K C P Y I I G T M N
	1110 1120 1130 1140 1150
1 1101..1150 UNC53_hu	CAGGCTACCT CTTCGACTCC CAACCTGCAG CTTCACCATA ACTTCAGATG
5 1099..1148	Q A T S S T P N L Q L H H N F R W
	NOVEL SEQUENCE NOT IN fig 11a
	1160 1170 1180 1190 1200
1 1151..1200 UNC53_hu	GGTGCTTTGT GCCAACCACA CGGAGCCTGT GAAGGGTTTC CTTGGCCGAT
5 1149..1198	V L C A N H T E P V K G F L G R
	1210 1220 1230 1240 1250
1 1201..1250 UNC53_hu	TCCTGAGGAG GAAGCTCATG GAAACAGAGA TCAGTGGGCG GGTGCGCAAT
5 1199..1248	F L R R K L M E T E I S G R V R N
	1260 1270 1280 1290 1300
1 1251..1300 UNC53_h	ATGGAGCTGG TAAAAATCAT TGACTGGATT CCCAAGGTCT GGCATCACCT
5 1249..1298	M E L V K I I D W I P K V W H H L
	1310 1320 1330 1340 1350
1 1301..1350 UNC53_h	CAACCGCTTC CTGGAGGCTC ACAGTTCCTC GGACGTCACC ATCGGCCCCC
5 1299..1348	N R F L E A H S S S D V T I G P
	1360 1370 1380 1390 1400
1 1351..1400 UNC53_h	CGCTCTTCCT GTCATGCCCC ATCGATGTGG ACGGCTCGAG AGTGTGGTTC
5 1349..1398	R L F L S C P I D V D G S R V W F

FIG. 11b CONTINUED.

```
                              1410      1420      1430      1440      1450
1 1401...1450 UNC53_hu ACCGACTTGT GGAACTATTC CATTATCCCC TATCTCCTGG AAGCCGTCAG
5 1399...1448           T  D  L   W  N  Y  S   I  I  F   Y  L  L   E  A  V  R

                              1460      1470      1480      1490      1500
1 1451...1500 UNC53_hu AGAAGGACTC CAGCTCTATG GAAGGCGCGC CCCCTGGGAG GATCCTGCCA
5 1449...1498            E  G  L   Q  L  Y   G  R  R  A   P  W  E   D  P  A

                              1510      1520      1530      1540      1550
1 1501...1550 UNC53_hu AGTGGGTGAT GGACACATAT CCATGGGCAG CCAGCCCACA ACAGCACGAG
5 1499...1548          K  W  V  M   D  T  Y   P  W  A   A  S  P  Q   Q  H  E

                              1560      1570      1580      1590      1600
1 1551...1600 UNC53_hu TGGCCTCCCC TGCTGCAGTT ACGGCCTGAG GATGTCGGCT TCGACGGCTA
5 1549...1598           W  P  P   L  L  Q  L   R  P  E   D  V  G   F  D  G  Y

                              1610      1620      1630      1640      1650
1 1601...1650 UNC53_hu CTCCATGCCT CGGGAGGGAT CGACAAGCAA GCAGATGCCC CCCAGTGATG
5 1599...1648            S  M  P   R  E  G   S  T  S  K   Q  M  P   P  S  D

                              1660      1670      1680      1690      1700
1 1651...1700 UNC53_hu CTGAAGGTGA CCCGCTGATG AACATGCTGA TGAGGCTGCA GGAGGCAGCC
5 1649...1698          A  E  G  D   P  L  M   N  M  L   M  R  L  Q   E  A  A

                              1710      1720      1730      1740      1750
1 1701...1750 UNC53_hu AACTACTCCA GCCCCCAGAG CTATGACAGC GACTCCAACA GCAACAGCCA
5 1699...1748           N  Y  S   S  P  Q  S   Y  D  S   D  S  N   S  N  S  H

                              1760      1770      1780      1790      1800
1 1751...1800 UNC53_hu TCACGATGAC ATCTTGGACT CCTCTTTGGA GTCCACTCTG TGACAGGGGC
5 1749...1790            H  D  D   I  L  D   S  S  L  E   S  T  L  *
                                                          end ORF
                              1810      1820      1830      1840      1850
1 1801...1850 UNC53_hu CCGGAGCCCA GCGCCCTCCT CTTCTCCTCA CCGCATTCCA CCTGCATCCC
5 ---,---               <==

                              1860      1870      1880      1890      1900
1 1851...1900 UNC53_hu CACATCACCC TGAAGATGAC TTCCTGAGCC AGCCCCAGCC ACAGCCTTAG
5 ---,---               <==

                              1910      1920      1930      1940      1950
1 1901...1941 UNC53_hu AGCTGCGGGA ACACCGAGAC CCCCGTCCTT CAGCCTCGAC T
5 ---,---               <==
```

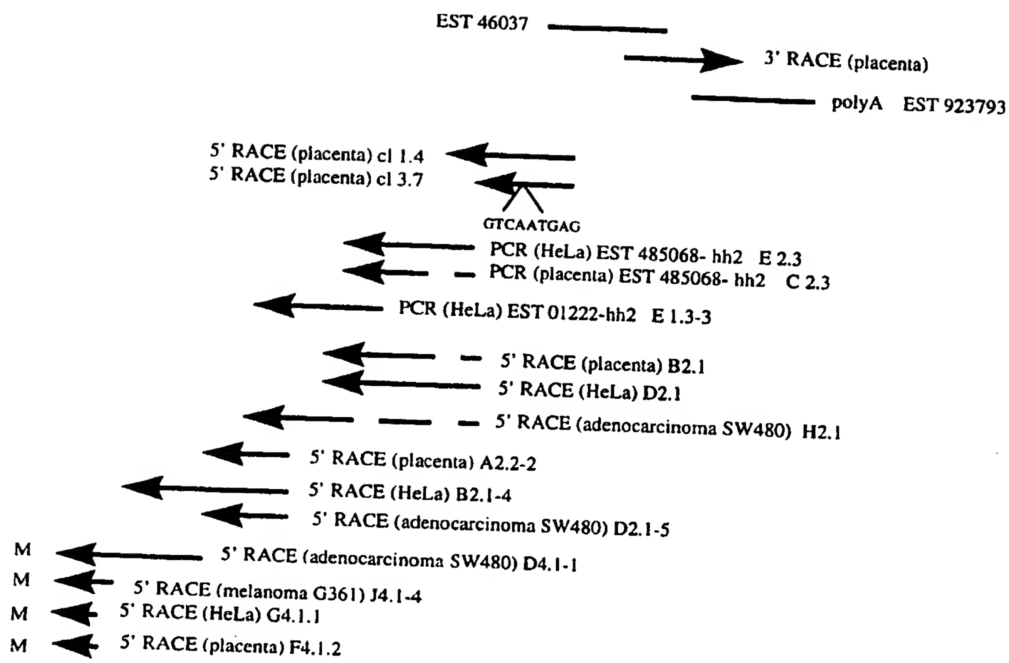

EST 46037
3' RACE (placenta)
polyA EST 923793
5' RACE (placenta) cl 1.4
5' RACE (placenta) cl 3.7
GTCAATGAG
PCR (HeLa) EST 485068- hh2 E 2.3
PCR (placenta) EST 485068- hh2 C 2.3
PCR (HeLa) EST 01222-hh2 E 1.3-3
5' RACE (placenta) B2.1
5' RACE (HeLa) D2.1
5' RACE (adenocarcinoma SW480) H2.1
5' RACE (placenta) A2.2-2
5' RACE (HeLa) B2.1-4
5' RACE (adenocarcinoma SW480) D2.1-5
M
5' RACE (adenocarcinoma SW480) D4.1-1
M
5' RACE (melanoma G361) J4.1-4
M
5' RACE (HeLa) G4.1.1
M
5' RACE (placenta) F4.1.2

Figure 11c

FIGURE 11d

11/25/97 02:17PM hh2UNC53_971120 Page 1

```
+1    I  Y  T  D  W  A  N  H  Y  L  T  K  S  G  H  K  R  L  I  K  D  L  Q  Q  D  V
1     ATCTACACAG ACTGGGCCAA TCATTACCTA ACCAAATCCG GCCACAAGCG TCTCATCAAG GATCTCCAGC AAGATGTGAC
      TAGATGTGTC TGACCCGGTT AGTAATGGAT TGGTTTAGGC CGGTGTTCGC AGAGTAGTTC CTAGAGGTCG TTCTACACTG
                                                         97080802

+1    D  G  V  L  L  A  Q  I  I  Q  V  V  A  N  E  K  I  E  D  I  N  G  C  P  K  N
81    AGATGGCGTC CTCCTGGCCC AGATTATCCA GGTTGTGGCA AATGAAAAGA TTGAAGACAT CAATGGCTGT CCGAAGAACA
      TCTACCGCAG GAGGACCGGG TCTAATAGGT CCAACACCGT TTACTTTTCT AACTTCTGTA GTTACCGACA GGCTTCTTGT
                 97080801

+1    R  S  Q  M  I  E  N  I  D  A  C  L  N  F  L  A  A  K  G  I  N  I  Q  G  L  S  A
161   GATCCCAAAT GATTGAAAAC ATAGATGCCT GCTTGAATTT CCTGGCAGCT AAGGGAATAA ACATCCAGGG GCTGTCTGCA
      CTAGGGTTTA CTAACTTTTG TATCTACGGA CGAACTTAAA GGACCGTCGA TTCCCTTATT TGTAGGTCCC CGACAGACGT

+1    E  E  I  R  N  G  N  L  K  A  I  L  G  L  F  F  S  L  S  R  Y  K  Q  Q  Q  Q
241   GAAGAGATCA GGAATGGAAA CCTCAAGGCC ATTCTAGGCC TCTTCTTCAG CCTCTCCCGA TACAAGCAGC AGCAGCAGCA
      CTTCTCTAGT CCTTACCTTT GGAGTTCCGG TAAGATCCGG AGAAGAAGTC GGAGAGGGCT ATGTTCGTCG TCGTCGTCGT

+1    P  Q  K  Q  H  L  S  S  P  L  P  P  A  V  S  Q  V  A  G  A  P  S  Q  C  Q  A
321   GCCCCAGAAG CAGCACCTCT CCTCACCTCT GCCGCCCGCC GTATCCCAGG TGGCCGGGGC CCCCTCCCAG TGCCAGGCTG
      CGGGGTCTTC GTCGTGGAGA GGAGTGGAGA CGGCGGGCGG CATAGGGTCC ACCGGCCCCG GGGGAGGGTC ACGGTCCGAC

+1    G  T  P  Q  Q  Q  V  P  V  T  P  Q  A  P  C  Q  P  H  Q  P  A  P  H  Q  Q  S  K
401   GCACCCCTCA GCAGCAGGTG CCAGTCACTC CCCAAGCCCC GTGCCAGCCT CACCAGCCAG CGCCACATCA GCAGTCAAAA
      CGTGGGGAGT CGTCGTCCAC GGTCAGTGAG GGGTTCGGGG CACGGTCGGA GTGGTCGGTC GCGGTGTAGT CGTCAGTTTT

+1    A  Q  A  E  M  Q  S  R  L  S  G  P  T  A  R  V  S  A  A  G  S  E  A  K  T  R
481   GCACAAGCTG AAATGCAGTC CAGACTTTCA GGTCCTACCG CGAGGGTATC CGCTGCAGGC AGCGAGGCCA AAACACGCGG
      CGTGTTCGAC TTTACGTCAG GTCTGAAAGT CCAGGATGGC GCTCCCATAG GCGACGTCCG TCGCTCCGGT TTTGTGCGCC
            97080804

+1    G  S  T  T  A  N  N  R  R  S  Q  S  F  N  N  Y  D  K  S  K  P  V  T  S  P  P
561   AGGGTCAACT ACTGCTAACA ACCGACGCAG CCAGAGCTTT AACAACTATG ATAAATCCAA ACCAGTCACC TCCCCACCCC
      TCCCAGTTGA TGACGATTGT TGGCTGCGTC GGTCTCGAAA TTGTTGATAC TATTTAGGTT TGGTCAGTGG AGGGGTGGGG
        97080803

+1    P  P  P  S  S  H  E  K  E  P  L  A  S  S  A  S  S  H  P  G  M  S  D  N  A  P  A
641   CACCGCCAAG CAGCCACGAG AAAGAGCCTT TGGCAAGTTC AGCCTCCTCC CACCCCGGAA TGAGTGACAA TGCACCTGCT
      GTGGCGGTTC GTCGGTGCTC TTTCTCGGAA ACCGTTCAAG TCGGAGGAGG GTGGGGCCTT ACTCACTGTT ACGTGGACGA

+1    S  L  E  S  G  S  S  S  T  P  T  N  C  S  T  S  S  A  I  P  Q  P  G  A  A  T
721   TCCTTGGAGA GCGGCAGCAG CTCCACCCCT ACTAATTGCA GTACCTCCTC GGCCATCCCG CAGCCCGGTG CAGCCACCAA
      AGGAACCTCT CGCCGTCGTC GAGGTGGGGA TGATTAACGT CATGGAGGAG CCGGTAGGGC GTCGGGCCAC GTCGGTGGTT

+1    P  W  R  S  K  S  L  S  V  K  H  S  A  T  V  S  M  L  S  V  K  P  P  G  P  E
801   GCCTTGGCGC AGCAAATCCC TCAGCGTGAA GCACAGTGCC ACGGTATCCA TGCTCTCGGT CAAGCCTCCT GGGCCTGAGG
      CGGAACCGCG TCGTTTAGGG AGTCGCACTT CGTGTCACGG TGCCATAGGT ACGAGAGCCA GTTCGGAGGA CCCGGACTCC

+1    A  P  R  P  T  P  E  A  M  K  P  A  P  N  N  Q  K  S  M  L  E  K  L  K  L  F  N
881   CCCCCAGGCC CACACCTGAA GCCATGAAGC CGGCCCCCAA CAATCAGAAG TCCATGCTGG AAAAGCTGAA ACTTTTCAAC
      GGGGGTCCGG GTGTGGACTT CGGTACTTCG GCCGGGGGTT GTTAGTCTTC AGGTACGACC TTTTCGACTT TGAAAAGTTG

+1    S  K  G  G  S  K  A  G  E  G  P  G  S  R  D  T  S  C  E  R  L  E  T  L  P  S
961   AGTAAAGGGG GCTCAAAGGC AGGTGAGGGG CCGGGGTCCC GGGACACAAG CTGTGAGCGG CTGGAGACTC TGCCCAGCTT
      TCATTTCCCC CGAGTTTCCG TCCACTCCCC GGCCCCAGGG CCCTGTGTTC GACACTCGCC GACCTCTGAG ACGGGTCGAA

+1    E  E  S  E  E  L  E  A  A  S  R  M  L  T  T  V  G  P  A  S  S  S  P  K  I  A
1041  CGAAGAGAGC GAGGAGCTGG AGGCCGCCAG TCGCATGCTC ACCACCGTGG GCCCTGCTTC CAGCAGCCCC AAGATTGCAC
      GCTTCTCTCG CTCCTCGACC TCCGGCGGTC AGCGTACGAG TGGTGGCACC CGGGACGAAG GTCGTCGGGG TTCTAACGTG

+1    L  K  G  I  A  Q  R  T  F  S  R  A  L  T  N  K  K  S  S  L  K  G  N  E  K  E  K
1121  TCAAGGGCAT TGCCCAGAGG ACTTTTAGCC GGGCACTGAC CAACAAGAAG AGTTCTCTGA AAGGCAATGA GAAAGAGAAG
      AGTTCCCGTA ACGGGTCTCC TGAAAATCGG CCCGTGACTG GTTGTTCTTC TCAAGAGACT TTCCGTTACT CTTTCTCTTC

+1    E  K  Q  Q  R  E  K  D  K  E  K  S  K  D  L  A  K  R  A  S  V  T  E  R  L  D
1201  GAGAAACAAC AGCGGGAGAA GGATAAGGAG AAAAGCAAGG ACCTTGCCAA GAGAGCCTCT GTGACGGAGA GGCTGGACCT
      CTCTTTGTTG TCGCCCTCTT CCTATTCCTC TTTTCGTTCC TGGAACGGTT CTCTCGGAGA CACTGCCTCT CCGACCTGGA
```

11/25/97 02:17PM hh2UNC53_971120 Page 2

```
+1    K  E  E  P  K  E  D  P  S  G  A  A  V  P  E  M  P  K  K  S  S  K  I  A  S  F
1281  CAAGGAGGAG CCAAAAGAAG ACCCCAGTGG AGCAGCTGTG CCCGAGATGC CAAAAAAGTC CTCCAAGATT GCCAGCTTCA
      GTTCCTCCTC GGTTTTCTTC TGGGGTCACC TCGTCGACAC GGGCTCTACG GTTTTTTCAG GAGGTTCTAA CGGTCGAAGT

+1    I  P  K  G  G  K  L  N  S  A  K  K  E  P  M  A  P  S  H  S  G  I  P  K  P  G  M
1361  TCCCCAAAGG GGGGAAGCTC AACAGTGCCA AGAAGGAGCC CATGGCCCCT TCCCACAGTG GAATACCAAA ACCAGGAATG
      AGGGGTTTCC CCCCTTCGAG TTGTCACGGT TCTTCCTCGG GTACCGGGGA AGGGTGTCAC CTTATGGTTT TGGTCCTTAC

+1    K  S  M  P  G  K  S  P  S  A  P  A  P  S  K  E  G  E  R  S  R  S  G  K  L  S
1441  AAGAGCATGC CCGGGAAATC CCCAAGTGCC CCAGCGCCTT CCAAGGAAGG GGAGCGGAGC CGGAGTGGGA AGCTGAGCTC
      TTCTCGTACG GGCCCTTTAG GGGTTCACGG GGTCGCGGAA GGTTCCTTCC CCTCGCCTCG GCCTCACCCT TCGACTCGAG

+1    G  L  P  Q  Q  K  P  Q  L  D  G  R  H  S  S  S  S  S  S  L  A  S  S  E  G  K
1521  AGGACTCCCC CAGCAGAAGC CCCAGCTGGA CGGCAGACAC TCCAGTTCCT CTTCCAGCCT GGCGTCCTCA GAAGGAAAAG
      TCCTGAGGGG GTCGTCTTCG GGGTCGACCT GCCGTCTGTG AGGTCAAGGA GAAGGTCGGA CCGCAGGAGT CTTCCTTTTC

+1    G  P  G  G  T  T  L  N  H  S  I  S  S  Q  T  V  S  G  S  V  G  T  T  Q  T  T  G
1601  GCCCAGGAGG GACCACCCTG AACCACAGCA TCAGCAGCCA GACTGTCAGT GGGTCTGTCG GGACCACCCA GACCACAGGA
      CGGGTCCTCC CTGGTGGGAC TTGGTGTCGT AGTCGTCGGT CTGACAGTCA CCCAGACAGC CCTGGTGGGT CTGGTGTCCT

+1    S  N  T  V  S  V  Q  L  P  Q  P  Q  Q  Q  Y  N  H  P  N  T  A  T  V  A  P  F
1681  AGCAATACCG TCAGTGTTCA GCTACCTCAG CCCCAGCAGC AATACAACCA TCCCAACACT GCCACGGTTG CACCTTTCCT
      TCGTTATGGC AGTCACAAGT CGATGGAGTC GGGGTCGTCG TTATGTTGGT AGGGTTGTGA CGGTGCCAAC GTGGAAAGGA

+1    Y  R  S  Q  T  D  T  E  G  N  V  T  A  E  S  S  S  T  G  V  S  V  E  P  S  H
1761  GTACAGGTCT CAGACGGACA CTGAAGGGAA TGTTACTGCC GAGTCAAGCT CAACAGGTGT GAGCGTGGAG CCCAGCCACT
      CATGTCCAGA GTCTGCCTGT GACTTCCCTT ACAATGACGG CTCAGTTCGA GTTGTCCACA CTCGCACCTC GGGTCGGTGA

+1    F  T  K  T  G  Q  P  A  L  E  E  L  T  G  E  D  P  E  A  R  R  L  R  T  V  K  N
1841  TCACCAAGAC TGGACAGCCT GCTCTGGAAG AACTCACTGG GGAAGATCCT GAGGCTCGGC GGCTGCGGAC AGTGAAGAAC
      AGTGGTTCTG ACCTGTCGGA CGAGACCTTC TTGAGTGACC CCTTCTAGGA CTCCGAGCCG CCGACGCCTG TCACTTCTTG

+1    I  A  D  L  R  Q  N  L  E  E  T  M  S  S  L  R  G  T  Q  V  T  H  S  T  L  E
1921  ATCGCTGATC TGCGGCAGAA TTTGGAGGAA ACCATGTCCA GTTTAAGGGG AACTCAGGTT ACACACAGCA CATTGGAAAC
      TAGCGACTAG ACGCCGTCTT AAACCTCCTT TGGTACAGGT CAAATTCCCC TTGAGTCCAA TGTGTGTCGT GTAACCTTTG
                                                    97041702

+1    T  F  D  T  N  V  T  T  E  M  S  G  R  S  I  L  S  L  T  G  R  P  T  P  L  S
2001  CACGTTTGAC ACCAATGTCA CCACGGAGAT GAGTGGCCGT AGCATACTCA GCTTGACAGG GAGGCCCACA CCTCTGTCCT
      GTGCAAACTG TGGTTACAGT GGTGCCTCTA CTCACCGGCA TCGTATGAGT CGAACTGTCC CTCCGGGTGT GGAGACAGGA
                                  97041701

+1    W  R  L  G  Q  S  S  P  R  L  Q  A  G  D  A  P  S  M  G  N  G  Y  P  P  R  A  N
2081  GGAGACTGGG CCAGTCCAGC CCTCGGCTCC AAGCAGGAGA CGCCCCCTCA ATGGGCAATG GGTATCCCCC TCGAGCCAAC
      CCTCTGACCC GGTCAGGTCG GGAGCCGAGG TTCGTCCTCT GCGGGGGAGT TACCCGTTAC CCATAGGGGG AGCTCGGTTG

+1    A  S  R  F  I  N  T  E  S  G  R  Y  V  Y  S  A  P  L  R  R  Q  L  A  S  R  G
2161  GCCAGCAGGT TCATCAACAC TGAGTCAGGT CGCTATGTGT ACTCCGCCCC TCTGAGAAGG CAGCTGGCCT CCCGGGGCAG
      CGGTCGTCCA AGTAGTTGTG ACTCAGTCCA GCGATACACA TGAGGCGGGG AGACTCTTCC GTCGACCGGA GGGCCCCGTC

+1    S  V  C  H  V  D  V  S  D  K  A  G  D  E  M  D  L  E  G  I  S  M  D  A  P  G
2241  TAGTGTCTGC CATGTGGACG TCTCAGACAA GGCAGGAGAT GAGATGGACC TGGAAGGCAT CAGCATGGAC GCCCCCGGCT
      ATCACAGACG GTACACCTGC AGAGTCTGTT CCGTCCTCTA CTCTACCTGG ACCTTCCGTA GTCGTACCTG CGGGGGCCGA

+1    Y  M  S  D  G  D  V  L  S  K  N  I  R  T  D  D  I  T  S  G  Y  M  T  D  G  G  L
2321  ACATGAGCGA TGGGGATGTT CTGAGCAAGA ACATCCGGAC CGATGACATT ACAAGCGGAT ACATGACTGA TGGTGGACTT
      TGTACTCGCT ACCCCTACAA GACTCGTTCT TGTAGGCCTG GCTACTGTAA TGTTCGCCTA TGTACTGACT ACCACCTGAA

+1    G  L  Y  T  R  R  L  N  R  L  P  D  G  M  A  V  V  R  E  T  L  Q  R  N  T  S
2401  GGCCTCTATA CCCGTCGCCT GAACCGGCTC CCTGATGGGA TGGCTGTGGT ACGGGAGACC CTGCAACGAA ATACCTCCCT
      CCGGAGATAT GGGCAGCGGA CTTGGCCGAG GGACTACCCT ACCGACACCA TGCCCTCTGG GACGTTGCTT TATGGAGGGA

+1    G  L  G  D  A  D  S  W  D  E  S  S  S  V  S  S  G  I  S  D  T  I  D  N  L  S
2481  GGGCCTCGGA GACGCTGACA GCTGGGACGA CAGCAGCTCC GTCAGCAGCG GCATCAGCGA CACCATAGAC AACCTCAGCA
      CCCGGAGCCT CTGCGACTGT CGACCCTGCT GTCGTCGAGG CAGTCGTCGC CGTAGTCGCT GTGGTATCTG TTGGAGTCGT
                               97012801
```

11/25/97 01:00PM hh2UNC53_971120 Page 3

```
   +1 T  D  D  I   N  T  S   S  S  I  S   S  Y  A   N  T  P  A   S  S  R   K  N  L   D  V  Q
2561  CTGATGACAT CAACACCAGC TCCTCCATCA GCTCTTATGC CAACACACCT GCCTCCTCTC GAAAAAACCT GGATGTGCAG
      GACTACTGTA GTTGTGGTCG AGGAGGTAGT CGAGAATACG GTTGTGTGGA CGGAGGAGAG CTTTTTTGGA CCTACACGTC
.........................................................................................
   +1 T  D  A  E   K  H  S   Q  V  E   R  N  S  L   W  S  G   D  D  V   K  K  S  D   G  G
2641  ACTGATGCTG AGAAGCACTC ACAGGTGGAG AGGAATTCCC TGTGGTCTGG TGATGATGTC AAGAAATCAG ACGGAGGCTC
      TGACTACGAC TCTTCGTGAG TGTCCACCTC TCCTTAAGGG ACACCAGACC ACTACTACAG TTCTTTAGTC TGCCTCCGAG
.........................................................................................
   +1 D  S  G   I  K  M  E   P  G  S   K  W  R   R  N  P  S   D  V  S   D  E  S   D  K  S
2721  AGACAGCGGC ATAAAAATGG AGCCAGGTTC CAAGTGGAGG CGGAATCCTT CTGATGTGTC TGACGAGTCC GACAAAAGCA
      TCTGTCGCCG TATTTTTACC TCGGTCCAAG GTTCACCTCC GCCTTAGGAA GACTACACAG ACTGCTCAGG CTGTTTTCGT
.........................................................................................
   +1 T  S  G  K   K  N  P   V  I  S  Q   T  G  S   W  R  R   G  M  T  A   Q  V  G   I  T  M
2801  CGTCGGGCAA GAAGAATCCT GTCATCTCCC AGACAGGCTC ATGGCGGCGA GGCATGACAG CTCAGGTGGG CATCACCATG
      GCAGCCCGTT CTTCTTAGGA CAGTAGAGGG TCTGTCCGAG TACCGCCGCT CCGTACTGTC GAGTCCACCC GTAGTGGTAC
.........................................................................................
   +1 P  R  T  K   P  S  A   P  A  G   A  L  K  T   P  G  T   G  K  T   D  D  A  K   V  S
2881  CCAAGGACGA AGCCTTCAGC CCCGGCAGGC GCACTGAAGA CCCCAGGAAC TGGAAAAACA GACGACGCAA AGGTGTCTGA
      GGTTCCTGCT TCGGAAGTCG GGGCCGTCCG CGTGACTTCT GGGGTCCTTG ACCTTTTTGT CTGCTGCGTT TCCACAGACT
.........................................................................................
   +1 K  G  R   L  S  P  K   A  S  Q   V  K  R   S  P  S  D   A  G  R   S  S  G   D  E  S
2961  GAAAGGAAGG CTTTCTCCTA AAGCCTCCCA GGTGAAGCGC TCCCCATCAG ATGCAGGCCG GAGCAGTGGT GACGAATCCA
      CTTTCCTTCC GAAAGAGGAT TTCGGAGGGT CCACTTCGCG AGGGGTAGTC TACGTCCGGC CTCGTCACCA CTGCTTAGGT
.........................................................................................
```

97012802

```
   +1 K  K  P  L   P  S  S   S  R  T  P   T  A  N   A  N  S   F  G  F  K   K  Q  S   G  S  A
3041  AAAAGCCCCT CCCCAGCAGC TCTAGGACAC CTACTGCCAA TGCCAACAGC TTTGGGTTCA AGAAGCAGAG TGGTTCCGCC
      TTTTCGGGGA GGGGTCGTCG AGATCCTGTG GATGACGGTT ACGGTTGTCG AAACCCAAGT TCTTCGTCTC ACCAAGGCGG
.........................................................................................
   +1 A  G  L  A   N  I  T   A  S  G   V  T  V  T   S  R  S   A  T  L   G  K  I  P   K  S
3121  GCCGGCCTGG CCAATGATCAC AGCCAGCGGG GTGACTGTCA CCAGCAGGTC AGCCACACTG GGCAAAATCC CAAAGTCATC
      CGGCCGGACC GGTACTAGTG TCGGTCGCCC CACTGACAGT GGTCGTCCAG TCGGTGTGAC CCGTTTTAGG GTTTCAGTAG
.........................................................................................
   +1 A  L  V   S  R  S  A   G  R  K   S  S  M   D  G  A  Q   N  Q  D   D  G  Y   L  A  L
3201  TGCACTCGTC AGTCGGTCTG CTGGTCGGAA GTCAAGTATG GATGGGGCTC AGAATCAGGA TGACGGGTAT CTAGCCCTAA
      ACGTGAGCAG TCAGCCAGAC GACCAGCCTT CAGTTCATAC CTACCCCGAG TCTTAGTCCT ACTGCCCATA GATCGGGATT
.........................................................................................
   +1 S  S  R  T   N  L  Q   Y  R  S  L   P  R  P   S  K  S   N  S  R  N   G  A  G   N  R  S
3281  GCTCCCGGAC AAACCTTCAG TACCGGAGTT TGCCGAGGCC CAGTAAGTCC AACAGCCGGA ACGGGGCTGG GAACAGGTCT
      CGAGGGCCTG TTTGGAAGTC ATGGCCTCAA ACGGCTCCGG GTCATTCAGG TTGTCGGCCT TGCCCCGACC CTTGTCCAGA
.........................................................................................
   +1 S  T  S  S   I  D  S   N  I  S   S  K  S  A   G  L  P   V  P  K   L  R  E  P   S  K
3361  AGCACCAGCA GCATAGATTC CAACATTAGC AGCAAGTCCG CAGGCCTGCC AGTGCCCAAA CTGAGGGAGC CTTCCAAAAC
      TCGTGGTCGT CGTATCTAAG GTTGTAATCG TCGTTCAGGC GTCCGGACGG TCACGGGTTT GACTCCCTCG GAAGGTTTTG
.........................................................................................
   +1 A  L  G   S  S  L  P   G  L  V   N  Q  T   D  K  E  K   G  I  S   S  D  N   E  S  V
3441  AGCCCTAGGC AGCTCTCTAC CAGGTCTGGT CAACCAAACA GATAAGGAGA AAGGCATCTC ATCAGACAAC GAGAGTGTGG
      TCGGGATCCG TCGAGAGATG GTCCAGACCA GTTGGTTTGT CTATTCCTCT TTCCGTAGAG TAGTCTGTTG CTCTCACACC
.........................................................................................
   +1 A  S  C  N   S  V  K   V  N  P  A   A  Q  P   V  S  S   P  A  Q  T   S  L  Q   P  G  A
3521  CTTCCTGTAA CTCGGTGAAA GTGAATCCGG CAGCCCAGCC TGTGTCCAGT CCGGCTCAGA CCAGTCTCCA GCCTGGAGCC
      GAAGGACATT GAGCCACTTT CACTTAGGCC GTCGGGTCGG ACACAGGTCA GGCCGAGTCT GGTCAGAGGT CGGACCTCGG
.........................................................................................
   +1 K  Y  P  D   V  A  S   P  T  L   R  R  L  F   G  G  K   P  T  K   Q  V  P  I   A  T
3601  AAGTACCCAG ATGTGGCCTC TCCCACACTC CGCAGACTCT TTGGTGGGAA GCCTACCAAG CAAGTGCCCA TCGCCACAGC
      TTCATGGGTC TACACCGGAG AGGGTGTGAG GCGTCTGAGA AACCACCCTT CGGATGGTTC GTTCACGGGT AGCGGTGTCG
.........................................................................................
   +1 E  N  M   K  N  S  V   V  I  S   N  P  H   A  T  M  T   Q  Q  G   N  L  D   S  P  S
3681  TGAAAACATG AAAAATTCGG TGGTCATCTC CAATCCTCAT GCCACCATGA CTCAGCAAGG TAACCTAGAC TCCCCGTCAG
      ACTTTTGTAC TTTTTAAGCC ACCAGTAGAG GTTAGGAGTA CGGTGGTACT GAGTCGTTCC ATTGGATCTG AGGGGCAGTC
.........................................................................................
   +1 G  S  G  V   L  S  S   G  S  S  S   P  L  Y   S  K  N   V  D  L  N   Q  S  P   L  A  S
3761  GCAGTGGCGT CCTGAGCAGT GGGAGCAGCA GTCCTCTCTA CAGCAAGAAT GTGGACCTCA ACCAGTCTCC GCTAGCCTCC
      CGTCACCGCA GGACTCGTCA CCCTCGTCGT CAGGAGAGAT GTCGTTCTTA CACCTGGAGT TGGTCAGAGG CGATCGGAGG
.........................................................................................
```

11/25/97 01:00PM hh2UNC53_971120 Page 4

```
   +1 S  P  S  S   A  H  S   A  P  S   N  S  L  T   W  G  T   N  A  S   S  S  S  A   V  S
3841  AGCCCCAGCT CAGCCCACTC GGCCCCTTCC AACAGCCTCA CCTGGGGCAC CAACGCCAGC AGCTCCTCCG CAGTTAGCAA
      TCGGGGTCGA GTCGGGTGAG CCGGGGAAGG TTGTCGGAGT GGACCCCGTG GTTGCGGTCG TCGAGGAGGC GTCAATCGTT
.........................................................................................
   +1 D  G  L   G  F  Q  S   V  S  S   L  H  T   S  C  E  S   I  D  I   S  L  S   S  G  G
3921  GGATGGCCTG GGCTTTCAGT CTGTCAGCAG CCTCCACACC AGCTGTGAGT CCATCGACAT CTCCCTCAGC AGTGGAGGGG
      CCTACCGGAC CCGAAAGTCA GACAGTCGTC GGAGGTGTGG TCGACACTCA GGTAGCTGTA GAGGGAGTCG TCACCTCCCC
.........................................................................................
   +1 V  P  S  H   N  S  S   T  G  L  I   A  S  S   K  D  D   S  L  T  P   F  V  R   T  N  S
4001  TCCCCAGCCA CAATTCTTCC ACTGGCCTCA TCGCCTCCTC CAAGGACGAC TCCTTGACTC CCTTTGTCAG AACTAACAGT
      AGGGGTCGGT GTTAAGAAGG TGACCGGAGT AGCGGAGGAG GTTCCTGCTG AGGAACTGAG GGAAACAGTC TTGATTGTCA
.........................................................................................
   +1 V  K  T  T   L  S  E   S  P  L   S  S  P  A   A  S  P   K  F  C   R  S  T  L   P  R
4081  GTGAAGACCA CACTGTCAGA AAGCCCTCTC TCTTCCCCTG CTGCTAGCCC TAAGTTCTGC AGAAGTACTC TGCCCAGGAA
      CACTTCTGGT GTGACAGTCT TTCGGGAGAG AGAAGGGGAC GACGATCGGG ATTCAAGACG TCTTCATGAG ACGGGTCCTT
.........................................................................................
   +1 Q  D  S   D  P  H  L   D  R  N   T  L  P   K  K  G  L   R  Y  T   P  T  S   Q  L  R
4161  ACAGGACAGT GACCCGCACC TTGATAGGAA CACTTTGCCT AAGAAAGGAC TCAGGTATAC TCCCACCTCC CAGCTTCGCA
      TGTCCTGTCA CTGGGCGTGG AACTATCCTT GTGAAACGGA TTCTTTCCTG AGTCCATATG AGGGTGGAGG GTCGAAGCGT
.........................................................................................
```

97010708

```
   +1 T  Q  E  D   A  K  E   W  L  R  S   H  S  A   G  G  L   Q  D  T  A   A  N  S   P  F  S
4241  CGCAAGAAGA TGCAAAAGAA TGGTTACGGT CCCATTCTGC AGGAGGCCTT CAGGACACCG CTGCCAATTC CCCCTTTTCC
      GCGTTCTTCT ACGTTTTCTT ACCAATGCCA GGGTAAGACG TCCTCCGGAA GTCCTGTGGC GACGGTTAAG GGGGAAAAGG
.........................................................................................
```

97010709

```
   +1 S  G  S  S   V  T  S   P  S  G   T  R  F  N   F  S  Q   L  A  S   P  T  T  V   T  Q
4321  TCTGGCTCCA GCGTGACTTC TCCCTCCGGA ACAAGATTCA ACTTTTCCCA GCTTGCGAGT CCCACCACTG TCACCCAGAT
      AGACCGAGGT CGCACTGAAG AGGGAGGCCT TGTTCTAAGT TGAAAAGGGT CGAACGCTCA GGGTGGTGAC AGTGGGTCTA
.........................................................................................
   +1 S  L  S   N  P  T  M   L  R  T   H  S  L   S  N  A  D   G  Q  Y   D  P  Y   T  D  S
4401  GAGCTTGTCC AACCCGACCA TGCTGAGGAC TCACAGCCTC TCCAATGCTG ATGGGCAGTA TGATCCATAC ACTGACAGCC
      CTCGAACAGG TTGGGCTGGT ACGACTCCTG AGTGTCGGAG AGGTTACGAC TACCCGTCAT ACTAGGTATG TGACTGTCGG
.........................................................................................
   +1 R  F  R  N   S  S  M   S  L  D  E   K  S  R   T  M  S   R  S  G  S   F  R  D   G  F  E
4481  GCTTCCGGAA TAGCTCCATG TCCCTGGATG AGAAGAGCAG AACCATGAGC CGTTCAGGCT CATTCCGGGA TGGGTTTGAA
      CGAAGGCCTT ATCGAGGTAC AGGGACCTAC TCTTCTCGTC TTGGTACTCG GCAAGTCCGA GTAAGGCCCT ACCCAAACTT
.........................................................................................
   +1 E  V  H  G   S  S  L   S  L  V   S  S  T  L   S  V  Y   S  T  P   E  E  K  C   Q  S
4561  GAAGTTCATG GATCCTCACT CTCCCTGGTT TCCAGCACAT TGTCAGTTTA TTCTACACCA GAAGAAAAAT GCCAGTCAGA
      CTTCAAGTAC CTAGGAGTGA GAGGGACCAA AGGTCGTGTA ACAGTCAAAT AAGATGTGGT CTTCTTTTTA CGGTCAGTCT
.........................................................................................
   +1 I  R  K   L  R  R  E   L  D  A   S  Q  E   K  V  S  A   L  T  T   Q  L  T   A  N  A
4641  GATTCGCAAG CTGCGGCGGG AACTGGATGC CTCCCAGGAG AAAGTTTCAG CTTTGACCAC CCAGCTGACA GCAAATGCTC
      CTAAGCGTTC GACGCCGCCC TTGACCTACG GAGGGTCCTC TTTCAAAGTC GAAACTGGTG GGTCGACTGT CGTTTACGAG
.........................................................................................
   +1 H  L  V  A   A  F  E   Q  S  L  G   N  M  T   I  R  L   Q  S  L  T   M  T  A   E  Q  K
4721  ACCTTGTGGC TGCCTTTGAA CAGAGTCTTG GTAACATGAC AATCAGGCTC CAGAGTCTGA CCATGACAGC TGAGCAGAAG
      TGGAACACCG ACGGAAACTT GTCTCAGAAC CATTGTACTG TTAGTCCGAG GTCTCAGACT GGTACTGTCG ACTCGTCTTC
.........................................................................................
   +1 D  S  E  L   N  E  L   R  K  T   I  E  L  L   K  K  Q   N  A  A   A  Q  A  A   I  N
4801  GATTCAGAAC TGAATGAGTT AAGAAAAACC ATTGAGCTGC TAAAGAAACA GAACGCAGCT GCCCAGGCTG CCATTAATGG
      CTAAGTCTTG ACTTACTCAA TTCTTTTTGG TAACTCGACG ATTTCTTTGT CTTGCGTCGA CGGGTCCGAC GGTAATTACC
.........................................................................................
   +1 V  I  N   T  P  E  L   N  C  K   G  N  G   T  A  Q  S   A  D  L   R  I  R   R  Q  H
4881  AGTAATTAAC ACACCTGAGC TCAACTGCAA AGGAAACGGC ACTGCCCAGT CTGCAGACCT CCGCATCCGC AGGCAGCACT
      TCATTAATTG TGTGGACTCG AGTTGACGTT TCCTTTGCCG TGACGGGTCA GACGTCTGGA GGCGTAGGCG TCCGTCGTGA
.........................................................................................
   +1 S  S  D  S   V  S  S   I  N  S  A   T  S  H   S  S  V   G  S  N  I   E  S  D   S  K  K
4961  CCTCAGACAG CGTCTCCAGC ATCAACAGTG CCACCAGCCA CTCCAGTGTG GGCAGCAACA TAGAGAGTGA CTCAAAGAAG
      GGAGTCTGTC GCAGAGGTCG TAGTTGTCAC GGTGGTCGGT GAGGTCACAC CCGTCGTTGT ATCTCTCACT GAGTTTCTTC
.........................................................................................
   +1 K  K  R  K   N  W  L   R  S  S   F  K  Q  A   F  G  K   K  K  S   P  K  S  A   S  S
5041  AAGAAGAGGA AGAACTGGTT ACGCAGCTCC TTCAAGCAAG CTTTCGGGAA GAAGAAGTCC CCAAAATCTG CGTCCTCTCA
      TTCTTCTCCT TCTTGACCAA TGCGTCGAGG AAGTTCGTTC GAAAGCCCTT CTTCTTCAGG GGTTTTAGAC GCAGGAGAGT
.........................................................................................
```

```
  +1  S  D  I   E  E  T  T   D  S  S   L  P  S   S  P  K  L   P  H  N   G  S  T   G  S  T
5121  TTCAGATATT GAGGAGACGA CGGATTCTTC TTTGCCTTCC TCACCAAAGT TACCGCACAA TGGGTCCACA GGTTCCACCC
      AAGTCTATAA CTCCTCTGCT GCCTAAGAAG AAACGGAAGG AGTGGTTTCA ATGGCGTGTT ACCCAGGTGT CCAAGGTGGG

  +1 P  L  L  R   N  S  H   S  N  S  L   I  S  E   C  H  D   S  E  A  E   T  V  M   Q  L  R
5201  CACTGCTGAG GAATTCTCAC TCCAACTCTC TAATTTCCGA ATGCATGGAT AGTGAAGCTG AGACCGTCAT GCAGCTCCGA
      GTGACGACTC CTTAAGAGTG AGGTTGAGAG ATTAAAGGCT TACGTACCTA TCACTTCGAC TCTGGCAGTA CGTCGAGGCT

  +1  N  E  L  R   D  K  E   M  K  L   T  D  I  R   L  E  A   L  S  S   A  H  Q  L   D  Q
5281  AATGAGTTAA GAGACAAGGA GATGAAGCTG ACGGATATCC GCTTAGAAGC TCTCAGTTCT GCCCACCAGC TGGACCAGCT
      TTACTCAATT CTCTGTTCCT CTACTTCGAC TGCCTATAGG CGAATCTTCG AGAGTCAAGA CGGGTGGTCG ACCTGGTCGA
                                                         EST 46037-R2

  +1  R  E  A   M  N  R  M   Q  S  E   I  E  K   L  K  A  E   N  D  R   L  K  S   E  S  Q
5361  CTGGGAGGCC ATGAACAGGA TGCAGAGTGA AATAGAGAAG CTGAAAGCTG AGAATGATCG GCTGAAGTCA GAGTCTCAAG
      GACCCTCCGG TACTTGTCCT ACGTCTCACT TTATCTCTTC GACTTTCGAC TCTTACTAGC CGACTTCAGT CTCAGAGTTC
                                                                      EST 46037-R1

  +1 G  S  G  C   S  R  A   P  S  Q  V   S  I  S   A  S  P   R  Q  S  M   G  L  S   Q  H  S
5441  GCAGTGGCTG CAGCCGGGCT CCTTCCCAAG TGTCCATCTC TGCCTCCCCG AGGCAGTCCA TGGGCCTCTC CCAGCACAGC
      CGTCACCGAC GTCGGCCCGA GGAAGGGTTC ACAGGTAGAG ACGGAGGGGC TCCGTCAGGT ACCCGGAGAG GGTCGTGTCG

  +1  L  N  L  T   E  S  T   S  L  D   M  L  L  D   D  T  G   E  C  S   A  R  K  E   G  G
5521  TTGAACCTCA CTGAGTCAAC CAGCCTGGAC ATGTTGCTGG ATGACACTGG TGAATGCTCG GCTCGGAAGG AAGGAGGCAG
      AACTTGGAGT GACTCAGTTG GTCGGACCTG TACAACGACC TACTGTGACC ACTTACGAGC CGAGCCTTCC TTCCTCCGTC

  +1  H  V  K   I  V  V  S   F  Q  E   E  M  K   W  K  E  D   S  R  P   H  L  F   L  I  G
5601  GCATGTTAAG ATAGTTGTCA GCTTTCAGGA GGAAATGAAG TGGAAGGAGG ATTCCAGACC ACACCTCTTT CTTATTGGCT
      CGTACAATTC TATCAACAGT CGAAAGTCCT CCTTTACTTC ACCTTCCTCC TAAGGTCTGG TGTGGAGAAA GAATAACCGA

  +1 C  I  G  V   S  G  K   T  K  W  D   V  L  D   G  V  V   R  R  L  F   K  E  Y   I  I  H
5681  GCATTGGAGT TAGTGGCAAG ACGAAGTGGG ATGTGCTCGA TGGGGTGGTT AGACGGCTGT TCAAAGAATA CATCATTCAT
      CGTAACCTCA ATCACCGTTC TGCTTCACCC TACACGAGCT ACCCCACCAA TCTGCCGACA AGTTTCTTAT GTAGTAAGTA

  +1  V  D  P  V   S  Q  L   G  L  N   S  D  S  V   L  G  Y   S  I  G   E  I  K  R   S  N
5761  GTCGACCCAG TGAGTCAGCT AGGGCTGAAT TCAGACAGCG TTCTTGGCTA CAGCATTGGA GAAATCAAGC GCAGCAACAC
      CAGCTGGGTC ACTCAGTCGA TCCCGACTTA AGTCTGTCGC AAGAACCGAT GTCGTAACCT CTTTAGTTCG CGTCGTTGTG
                                 EST 46037

  +1  S  E  T   P  E  L  L   P  C  G   Y  L  V   G  E  N  T   T  I  S   V  T  V   K  G  L
5841  TTCCGAAACA CCGGAGCTGC TTCCTTGTGG CTATCTGGTT GGAGAGAACA CGACCATCTC AGTGACTGTG AAAGGGCTCG
      AAGGCTTTGT GGCCTCGACG AAGGAACACC GATAGACCAA CCTCTCTTGT GCTGGTAGAG TCACTGACAC TTTCCCGAGC

  +1 A  E  N  S   L  D  S   L  V  F  E   S  L  I   P  K  P   I  L  Q  R   Y  V  S   L  L  I
5921  CAGAAAACAG CCTGGACTCA CTGGTGTTTG AGTCCTTGAT TCCCAAGCCC ATCCTGCAGC GCTACGTCTC CCTCCTGATA
      GTCTTTTGTC GGACCTGAGT GACCACAAAC TCAGGAACTA AGGGTTCGGG TAGGACGTCG CGATGCAGAG GGAGGACTAT

  +1  E  H  R  R   I  I  L   S  G  P   S  G  T  G   K  T  Y   L  A  N   K  L  S  E   Y  I
6001  GAGCACCGTC GGATCATTCT CTCTGGCCCC AGCGGCACTG GGAAAACCTA CCTGGCCAAC CGGCTGTCTG AGTATATAGT
      CTCGTGGCAG CCTAGTAAGA GAGACCGGGG TCGCCGTGAC CCTTTTGGAT GGACCGGTTG GCCGACAGAC TCATATATCA

  +1  L  R  E   G  R  E  L   T  D  G   V  I  A   T  F  N  V   D  H  K   S  S  K   E  L  R
6081  GCTTCGAGAG GGACGGGAGT TGACAGACGG GGTTATCGCC ACCTTTAACG TGGACCATAA GTCCAGCAAG GAATTGCGCC
      CGAAGCTCTC CCTGCCCTCA ACTGTCTGCC CCAATAGCGG TGGAAATTGC ACCTGGTATT CAGGTCGTTC CTTAACGCGG

  +1 Q  Y  L  S   N  L  A   D  Q  C  N   S  E  N   N  A  V   D  M  P  L   V  I  I   L  D  N
6161  AGTACCTGTC CAACCTTGCT GACCAGTGCA ACAGTGAGAA CAATGCTGTG GACATGCCCC TCGTCATCAT CCTGGACAAC
      TCATGGACAG GTTGGAACGA CTGGTCACGT TGTCACTCTT GTTACGACAC CTGTACGGGG AGCAGTAGTA GGACCTGTTG
                                      EST 46037.F1

  +1  L  H  H  V   S  S  L   G  E  I   F  N  G  L   L  N  C   K  Y  H   K  C  P  Y   I  I
6241  CTACACCACG TGAGCTCTCT GGGCGAGATC TTCAATGGGC TGCTCAACTG CAAGTACCAC AAATGCCCTT ACATAATTGG
      GATGTGGTGC ACTCGAGAGA CCCGCTCTAG AAGTTACCCG ACGAGTTGAC GTTCATGGTG TTTACGGGAA TGTATTAACC
                                                EST 46037.F2

  +1  T  M  N   Q  A  T  S   S  T  P   N  L  Q   L  H  H  N   F  R  W   V  L  C   A  N  H
6321  CACAATGAAC CAGGCTACCT CTTCGACTCC CAACCTGCAG CTTCACCATA ACTTCAGATG GGTGCTTTGT GCCAACCACA
      GTGTTACTTG GTCCGATGGA GAAGCTGAGG GTTGGACGTC GAAGTGGTAT TGAAGTCTAC CCACGAAACA CGGTTGGTGT
```



```
  +1 T  E  P  V   K  G  F   L  G  R  F   L  R  R   K  L  M   E  T  E  I   S  G  R   V  R  N
6401  CGGAGCCTGT GAAGGGTTTC CTTGGCCGAT TCCTGAGGAG GAAGCTCATG GAAACAGAGA TCAGTGGGCG GGTGCGCAAT
      GCCTCGGACA CTTCCCAAAG GAACCGGCTA AGGACTCCTC CTTCGAGTAC CTTTGTCTCT AGTCACCCGC CCACGCGTTA

  +1  M  E  L  V   K  I  I   D  W  I   P  K  V  W   H  H  L   N  R  F   L  E  A  H   S  S
6481  ATGGAGCTGG TAAAAATCAT TGACTGGATT CCCAAGGTCT GGCATCACCT CAACCGCTTC CTGGAGGCTC ACAGTTCCTC
      TACCTCGACC ATTTTTAGTA ACTGACCTAA GGGTTCCAGA CCGTAGTGGA GTTGGCGAAG GACCTCCGAG TGTCAAGGAG

  +1  D  V  T   I  G  P  R   L  F  L   S  C  P   I  D  V  D   G  S  R   V  W  F   T  D  L
6561  GGACGTCACC ATCGGCCCCC GGCTCTTCCT GTCATGCCCC ATCGATGTGG ACGGCTCGAG AGTGTGGTTC ACCGACTTGT
      CCTGCAGTGG TAGCCGGGGG CCGAGAAGGA CAGTACGGGG TAGCTACACC TGCCGAGCTC TCACACCAAG TGGCTGAACA

  +1 W  N  Y  S   I  I  P   Y  L  L  E   A  V  R   E  G  L   Q  L  Y  G   R  R  A   P  W  E
6641  GGAACTATTC CATTATCCCC TATCTCCTGG AAGCCGTCAG AGAAGGACTC CAGCTCTATG GAAGGCGCGC CCCCTGGGAG
      CCTTGATAAG GTAATAGGGG ATAGAGGACC TTCGGCAGTC TCTTCCTGAG GTCGAGATAC CTTCCGCGCG GGGGACCCTC

  +1  D  P  A  K   W  V  M   D  T  Y   P  W  A  A   S  P  Q   Q  H  E   W  P  P  L   L  Q
6721  GATCCTGCCA AGTGGGTGAT GGACACATAT CCATGGGCAG CCAGCCCACA ACAGCACGAG TGGCCTCCCC TGCTGCAGTT
      CTAGGACGGT TCACCCACTA CCTGTGTATA GGTACCCGTC GGTCGGGTGT TGTCGTGCTC ACCGGAGGGG ACGACGTCAA

  +1  R  P  E   D  V  G  F   D  G  Y   S  M  P   R  E  G  S   T  S  K   Q  M  P   P  S  D
6801  ACGGCCTGAG GATGTCGGCT TCGACGGCTA CTCCATGCCT CGGGAGGGAT CGACAAGCAA GCAGATGCCC CCCAGTGATG
      TGCCGGACTC CTACAGCCGA AGCTGCCGAT GAGGTACGGA GCCCTCCCTA GCTGTTCGTT CGTCTACGGG GGGTCACTAC

  +1 A  E  G  D   P  L  M   N  M  L  M   R  L  Q   E  A  A   N  Y  S  S   P  Q  S   Y  D  S
6881  CTGAAGGTGA CCCGCTGATG AACATGCTGA TGAGGCTGCA GGAGGCAGCC AACTACTCCA GCCCCCAGAG CTATGACAGC
      GACTTCCACT GGGCGACTAC TTGTACGACT ACTCCGACGT CCTCCGTCGG TTGATGAGGT CGGGGGTCTC GATACTGTCG
                                   EST 923793

  +1  D  S  N  S   N  S  H   H  E  D   I  L  D  S   S  L  E   S  T  L   *  Q  G  P   G  A
6961  GACTCCAACA GCAACAGCCA TCACGAAGAC ATCTTGGACT CCTCTTTGGA GTCCACTCTG TGACAGGGGC CCGGAGCCCA
      CTGAGGTTGT CGTTGTCGGT AGTGCTTCTG TAGAACCTGA GGAGAAACCT CAGGTGAGAC ACTGTCCCCG GGCCTCGGGT

  +1  R  P  P   L  L  L  T   A  F  H   L  H  P   P  H  H  P   E  D  D   F  L  S   Q  P  P
7041  GCGCCCTCCT CTTCTCCTCA CCGCATTCCA CCTGCATCCC CCACATCACC CTGAAGATGA CTTCCTGAGC CAGCCCCCAG
      CGCGGGAGGA GAAGAGGAGT GGCGTAAGGT GGACGTAGGG GGTGTAGTGG GACTTCTACT GAAGGACTCG GTCGGGGGTC

  +1 A  T  A  L   E  L  R   E  H  R  D   P  P  S   F  S  L   D  L  G  A   G  I  P   G  Q  L
7121  CCACAGCCTT AGAGCTGCGG GAACACCGAG ACCCCCCGTC CTTCAGCCTC GACCTGGGTG CAGGCATCCC GGGCCAGCTG
      GGTGTCGGAA TCTCGACGCC CTTGTGGCTC TGGGGGGCAG GAAGTCGGAG CTGGACCCAC GTCCGTAGGG CCCGGTCGAC

  +1  P  A  D  R   F  L  P   Q  R  E   L  H  Y  L   L  L  Y   F  N  Y   C  F  A  L   L  L
7201  CCTGCGGACC GCTTCCTTCC ACAGCGAGAA CTGCACTACC TTCTGTTGTA CTTTAATTAT TGTTTTGCCT TGTTGCTGTG
      GGACGCCTGG CGAAGGAAGG TGTCGCTCTT GACGTGATGG AAGACAACAT GAAATTAATA ACAAAACGGA ACAACGACAC

  +1  P  P  *   D  T  E  D   T  S  R   E  R  I   I  A  V  E   M  K  K   K  K  K   K  K  K
7281  ACCTCCCTAA GACACTGAAG ATACTTCTCG GGAAAGGATC ATCGCCGTTG AAATGAAAAA AAAAAAAAAA AAAAAAAAAA
      TGGAGGGATT CTGTGACTTC TATGAAGAGC CCTTTCCTAG TAGCGGCAAC TTTACTTTTT TTTTTTTTTT TTTTTTTTTT

  +1 K  K  K  K
7361  AAAAAAAAAA AAA
      TTTTTTTTTT TTT
```

11/24/97 02:33PM Fig 11e EST 923793 Page 1

```
+1  G  T  R  V  T  I  G   P  R  L   F  L  S  C   P  I  D   V  D  G   S  R  V  W   F  T  D
1   GGCACGAGGG TTACCATCGG CCCCCGGCTC TTCCTGTCAT GCCCCATCGA TGTGGACGGC TCGAGAGTGT GGTTCACCGA
    CCGTGCTCCC AATGGTAGCC GGGGGCCGAG AAGGACAGTA CGGGGTAGCT ACACCTGCCG AGCTCTCACA CCAAGTGGCT

+1   L  W  N   Y  S  I  I   P  Y  L   L  E  A   V  R  E  G   L  Q  L   Y  G  R   R  A  P
81  CTTGTGGAAC TATTCCATTA TCCCCTATCT CCTGGAAGCC GTCAGAGAAG GACTCCAGCT CTATGGAAGG CGCGCCCCCT
    GAACACCTTG ATAAGGTAAT AGGGGATAGA GGACCTTCGG CAGTCTCTTC CTGAGGTCGA GATACCTTCC GCGCGGGGGA

+1 W  E  D  P   A  K  W   V  M  D  T   Y  P  W   A  A  S   P  Q  Q  H   E  W  P   P  L  L
                                          NcoI
                                        ~~~~~~~
161 GGGAGGATCC TGCCAAGTGG GTGATGGACA CATATCCATG GGCAGCCAGC CCACAACAGC ACGAGTGGCC TCCCCTGCTG
    CCCTCCTAGG ACGGTTCACC CACTACCTGT GTATAGGTAC CCGTCGGTCG GGTGTTGTCG TGCTCACCGG AGGGGACGAC

+1  Q  L  R  P   E  D  V   G  F  D   G  Y  S  M   P  R  E   G  S  T   S  K  Q  M   P  P  S
241 CAGTTACGGC CTGAGGATGT CGGCTTCGAC GGCTACTCCA TGCCTCGGGA GGGATCGACA AGCAAGCAGA TGCCCCCCAG
    GTCAATGCCG GACTCCTACA GCCGAAGCTG CCGATGAGGT ACGGAGCCCT CCCTAGCTGT TCGTTCGTCT ACGGGGGGTC

+1   D  A  E   G  D  P  L   M  N  M   L  M  R   L  Q  E  A   A  N  Y   S  S  P   Q  S  Y
321 TGATGCTGAA GGTGACCCGC TGATGAACAT GCTGATGAGG CTGCAGGAGG CAGCCAACTA CTCCAGCCCC CAGAGCTATG
    ACTACGACTT CCACTGGGCG ACTACTTGTA CGACTACTCC GACGTCCTCC GTCGGTTGAT GAGGTCGGGG GTCTCGATAC

+1 D  S  D  S   N  S  N   S  H  H  E   D  I  L   D  S  S   L  E  S  T   L  *  Q   G  P  G
401 ACAGCGACTC CAACAGCAAC AGCCATCACG AAGACATCTT GGACTCCTCT TTGGAGTCCA CTCTGTGACA GGGGCCCGGA
    TGTCGCTGAG GTTGTCGTTG TCGGTAGTGC TTCTGTAGAA CCTGAGGAGA AACCTCAGGT GAGACACTGT CCCCGGGCCT

+1  A  Q  R  P   P  L  L   L  T  A   F  H  L  H   P  P  H   H  P  E   D  D  F  L   S  Q  P
481 GCCCAGCGCC CTCCTCTTCT CCTCACCGCA TTCCACCTGC ATCCCCCACA TCACCCTGAA GATGACTTCC TGAGCCAGCC
    CGGGTCGCGG GAGGAGAAGA GGAGTGGCGT AAGGTGGACG TAGGGGGTGT AGTGGGACTT CTACTGAAGG ACTCGGTCGG

+1   P  A  T   A  L  E  L   R  E  H   R  D  P   P  S  F  S   L  D  L   G  A  G   I  P  G
561 CCCAGCCACA GCCTTAGAGC TGCGGGAACA CCGAGACCCC CCGTCCTTCA GCCTCGACCT GGGTGCAGGC ATCCCGGGCC
    GGGTCGGTGT CGGAATCTCG ACGCCCTTGT GGCTCTGGGG GGCAGGAAGT CGGAGCTGGA CCCACGTCCG TAGGGCCCGG

+1 Q  L  P  A   D  R  F   L  P  Q  R   E  L  H   Y  L  L   L  Y  F  N   Y  C  F   A  L  L
641 AGCTGCCTGC GGACCGCTTC CTTCCACAGC GAGAACTGCA CTACCTTCTG TTGTACTTTA ATTATTGTTT TGCCTTGTTG
    TCGACGGACG CCTGGCGAAG GAAGGTGTCG CTCTTGACGT GATGGAAGAC AACATGAAAT TAATAACAAA ACGGAACAAC

+1  L  *  P  P   *  D  T   E  D  T   S  R  E  R   I  I  A   V  E  M   K  K  K  K   K  K  K
721 CTGTGACCTC CCTAAGACAC TGAAGATACT TCTCGGGAAA GGATCATCGC CGTTGAAATG AAAAAAAAAA AAAAAAAAAA
    GACACTGGAG GGATTCTGTG ACTTCTATGA AGAGCCCTTT CCTAGTAGCG GCAACTTTAC TTTTTTTTTT TTTTTTTTTT

+1   K  K  K   K  K  N  E   G  G  R   K  L
801 AAAAAAAAAA AAAAAAAACG AAGGCGGCCG CAAGCTT
    TTTTTTTTTT TTTTTTTTGC TTCCGCCGGC GTTCGAA
```

11/24/97 02:40PM hh2UNC53_5'Race_cl1.4_CON_REV.seq Page 1

```
 +1  I  P  L  T   I  G  L   E  R  P   P  R  Q  V   P  H  L   D  R  N   T  L  P  K   K  G  L
   1 ATCCCACTCA CTATAGGGCT CGAGCGGCCG CCCAGGCAGG TCCCGCACCT TGATAGGAAC ACTTTGCCTA AGAAAGGACT
     TAGGGTGAGT GATATCCCGA GCTCGCCGGC GGGTCCGTCC AGGGCGTGGA ACTATCCTTG TGAAACGGAT TCTTTCCTGA

 +1   R  Y  T   P  T  S  Q   L  R  T   Q  E  D   A  K  E  W   L  R  S   H  S  A   G  G  L
  81 CAGGTATACT CCCACCTCCC AGCTTCGCAC GCAAGAAGAT GCAAAAGAAT GGTTACGGTC CCATTCTGCA GGAGGCCTTC
     GTCCATATGA GGGTGGAGGG TCGAAGCGTG CGTTCTTCTA CGTTTTCTTA CCAATGCCAG GGTAAGACGT CCTCCGGAAG

 +1 Q  D  T  A   A  N  S   P  F  S  S   G  S  S   V  T  S   P  S  G  T   R  F  N   F  S  Q
 161 AGGACACCGC TGCCAATTCC CCCTTTTCCT CTGGCTCCAG CGTGACTTCT CCCTCCGGAA CAAGATTCAA CTTTTCCCAG
     TCCTGTGGCG ACGGTTAAGG GGGAAAAGGA GACCGAGGTC GCACTGAAGA GGGAGGCCTT GTTCTAAGTT GAAAAGGGTC

 +1  L  A  S  P   T  T  V   T  Q  M   S  L  S  N   P  T  M   L  R  T   H  S  L  S   N  A  D
 241 CTTGCGAGTC CCACCACTGT CACCCAGATG AGCTTGTCCA ACCCGACCAT GCTGAGGACT CACAGCCTCT CCAATGCTGA
     GAACGCTCAG GGTGGTGACA GTGGGTCTAC TCGAACAGGT TGGGCTGGTA CGACTCCTGA GTGTCGGAGA GGTTACGACT

 +1   G  Q  Y   D  P  Y  T   D  S  R   F  R  N   S  S  M  S   L  D  E   K  S  R   T  M  S
 321 TGGGCAGTAT GATCCATACA CTGACAGCCG CTTCCGGAAT AGCTCCATGT CCCTGGATGA GAAGAGCAGA ACCATGAGCC
     ACCCGTCATA CTAGGTATGT GACTGTCGGC GAAGGCCTTA TCGAGGTACA GGGACCTACT CTTCTCGTCT TGGTACTCGG

 +1 R  S  G  S   F  R  D   G  F  E  E   V  H  G   S  S  L   S  L  V  S   S  T  L   S  V  Y
 401 GTTCAGGCTC ATTCCGGGAT GGGTTTGAAG AAGTTCATGG ATCCTCACTC TCCCTGGTTT CCAGCACATT GTCAGTTTAT
     CAAGTCCGAG TAAGGCCCTA CCCAAACTTC TTCAAGTACC TAGGAGTGAG AGGGACCAAA GGTCGTGTAA CAGTCAAATA

 +1  S  T  P  E   E  K  C   Q  S  E   I  R  K  L   R  R  E   L  D  A   S  Q  E  K   V  S  A
 481 TCTACACCAG AAGAAAAATG CCAGTCAGAG ATTCGCAAGC TGCGGCGGGA ACTGGATGCC TCCCAGGAGA AAGTTTCAGC
     AGATGTGGTC TTCTTTTTAC GGTCAGTCTC TAAGCGTTCG ACGCCGCCCT TGACCTACGG AGGGTCCTCT TTCAAAGTCG

 +1   L  T  T   Q  L  T  A   N  A  H   L  V  A   A  F  E  Q   S  L  G   N  M  T   I  R  L
 561 TTTGACCACC CAGCTGACAG CAAATGCTCA CCTTGTGGCT GCCTTTGAAC AGAGTCTTGG TAACATGACA ATCAGGCTCC
     AAACTGGTGG GTCGACTGTC GTTTACGAGT GGAACACCGA CGGAAACTTG TCTCAGAACC ATTGTACTGT TAGTCCGAGG

 +1 Q  S  L  T   M  T  A   E  Q  K  D   S  E  L   N  E  L   R  K  T  I   E  L  L   K  K  Q
 641 AGAGTCTGAC CATGACAGCT GAGCAGAAGG ATTCAGAACT GAATGAGTTA AGAAAAACCA TTGAGCTGCT AAAGAAACAG
     TCTCAGACTG GTACTGTCGA CTCGTCTTCC TAAGTCTTGA CTTACTCAAT TCTTTTTGGT AACTCGACGA TTTCTTTGTC

 +1  N  A  A  A   Q  A  A   I  N  G   V  I  N  T   P  E  L   N  C  K   G  N  G  T   A  Q  S
 721 AACGCAGCTG CCCAGGCTGC CATTAATGGA GTAATTAACA CACCTGAGCT CAACTGCAAA GGAAACGGCA CTGCCCAGTC
     TTGCGTCGAC GGGTCCGACG GTAATTACCT CATTAATTGT GTGGACTCGA GTTGACGTTT CCTTTGCCGT GACGGGTCAG

 +1   A  D  L   R  I  R  R   Q  H  S   S  D  S   V  S  S  I   N  S  A   T  S  H   S  S  V
 801 TGCAGACCTC CGCATCCGCA GGCAGCACTC CTCAGACAGC GTCTCCAGCA TCAACAGTGC CACCAGCCAC TCCAGTGTGG
     ACGTCTGGAG GCGTAGGCGT CCGTCGTGAG GAGTCTGTCG CAGAGGTCGT AGTTGTCACG GTGGTCGGTG AGGTCACACC

 +1 G  S  N  I   E  S  D   S  K  K  K   K  R  K   N  W  L   R  S  S  F   K  Q  A   F  G  K
 881 GCAGCAACAT AGAGAGTGAC TCAAAGAAGA AGAAGAGGAA GAACTGGTTA CGCAGCTCCT TCAAGCAAGC TTTCGGGAAG
     CGTCGTTGTA TCTCTCACTG AGTTTCTTCT TCTTCTCCTT CTTGACCAAT GCGTCGAGGA AGTTCGTTCG AAAGCCCTTC

 +1  K  K  S  P   K  S  A   S  S  H   S  D  I  E   E  T  T   D  S  S   L  P  S  S   P  K  L
 961 AAGAAGTCCC CAAAATCTGC GTCCTCTCAT TCAGATATTG AGGAGACGAC GGATTCTTCT TTGCCTTCCT CACCAAAGTT
     TTCTTCAGGG GTTTTAGACG CAGGAGAGTA AGTCTATAAC TCCTCTGCTG CCTAAGAAGA AACGGAAGGA GTGGTTTCAA

 +1   P  H  N   G  S  T  G   S  T  P   L  L  R   N  S  H  S   N  S  L   I  S  E   C  M  D
1041 ACCGCACAAT GGGTCCACAG GTTCCACCCC ACTGCTGAGG AATTCTCACT CCAACTCTCT AATTTCCGAA TGCATGGATA
     TGGCGTGTTA CCCAGGTGTC CAAGGTGGGG TGACGACTCC TTAAGAGTGA GGTTGAGAGA TTAAAGGCTT ACGTACCTAT

 +1 S  E  A  E   T  V  M   Q  L  R  N   E  L  R   D  K  E   M  K  L  T   D  I  R
1121 GTGAAGCTGA GACCGTCATG CAGCTCCGAA ATGAGTTAAG AGACAAGGAG ATGAAGCTGA CGGATATCCG C
     CACTTCGACT CTGGCAGTAC GTCGAGGCTT TACTCAATTC TCTGTTCCTC TACTTCGACT GCCTATAGGC G
```

11/24/97 02:46PM hh2UNC53_5'Race_cl3.7_CON.seq Page 1

```
+2    E  F  E   L  G  T  L   T  I  G   L  E  R   P  P  G  Q   V  R  D   G  F  E   E  V  H
1     CGAATTCGAG CTCGGTACAC TCACTATAGG GCTCGAGCGG CCGCCCGGGC AGGTCGGGA TGGGTTTGAA GAAGTTCATG
      GCTTAAGCTC GAGCCATGTG AGTGATATCC CGAGCTCGCC GGCGGGCCCG TCCAGCCCT ACCCAAACTT CTTCAAGTAC

+2  G  S  S  L   S  L  V   S  S  T  S   S  V  Y   S  T  P   E  E  K  C   Q  S  E   I  R  K
81    GATCCTCACT CTCCTTGGTT TCCAGCACAT CGTCAGTTTA TTCTACACCA GAAGAAAAAT GCCAGTCAGA GATTCGCAAG
      CTAGGAGTGA GAGGAACCAA AGGTCGTGTA GCAGTCAAAT AAGATGTGGT CTTCTTTTTA CGGTCAGTCT CTAAGCGTTC

+2    L  R  R  E   L  D  A   S  Q  E   K  V  S  A   L  T  T   Q  L  T   A  N  A  H   L  V  A
161   CTGCGGCGGG AACTGGATGC CTCCCAGGAG AAAGTTTCAG CTTTGACCAC CCAGCTGACA GCAAATGCTC ACCTTGTGGC
      GACGCCGCCC TTGACCTACG GAGGGTCCTC TTTCAAAGTC GAAACTGGTG GGTCGACTGT CGTTTACGAG TGGAACACCG

+2    A  F  E   Q  S  L  G   N  M  T   I  R  L   Q  S  L  T   M  T  A   E  Q  K   D  S  E
241   AGCCTTTGAA CAGAGTCTTG GTAACATGAC AATCAGGCTC CAGAGTCTGA CCATGACAGC TGAGCAGAAG GACTCAGAAC
      TCGGAAACTT GTCTCAGAAC CATTGTACTG TTAGTCCGAG GTCTCAGACT GGTACTGTCG ACTCGTCTTC CTGAGTCTTG

+2  L  N  E  L   R  K  T   I  E  L  L   K  K  Q   N  A  A   A  Q  A  A   I  N  G   V  I  N
321   TGAATGAGTT AAGAAAAACC ATTGAGCTGC TAAAGAAACA GAACGCAGCT GCCCAGGCTG CCATTAATGG AGTAATTAAC
      ACTTACTCAA TTCTTTTTGG TAACTCGACG ATTTCTTTGT CTTGCGTCGA CGGGTCCGAC GGTAATTACC TCATTAATTG

+2    T  P  E  L   N  C  K   G  N  G   T  A  Q  S   A  D  L   R  I  R   R  Q  H  S   S  D  S
401   ACACCTGAGC TCAACTGCAA AGGAAACGGC ACTGCCCAGT CTGCAGACCT CCGCATCCGC AGGCAGCACT CCTCAGACAG
      TGTGGACTCG AGTTGACGTT TCCTTTGCCG TGACGGGTCA GACGTCTGGA GGCGTAGGCG TCCGTCGTGA GGAGTCTGTC

+2    V  S  S   I  N  S  A   T  S  H   S  S  V   G  S  N  I   E  S  D   S  K  K   K  K  R
481   CGTCTCCAGC ATCAACAGTG CCACCAGCCA CTCCAGCGTG GGCAGCAACA TAGAGAGTGA CTCAAAGAAG AAGAAGCGGA
      GCAGAGGTCG TAGTTGTCAC GGTGGTCGGT GAGGTCGCAC CCGTCGTTGT ATCTCTCACT GAGTTTCTTC TTCTTCGCCT

+2  K  N  W  V   N  E  L   R  S  S  F   K  Q  A   F  G  K   K  K  S  P   K  S  A   S  S  H
561   AGAACTGGGT CAATGAGTTA CGCAGCTCCT TCAAGCAAGC TTTCGGGAAG AAGAAGTCCC CAAAATCTGC GTCCTCTCAT
      TCTTGACCCA GTTACTCAAT GCGTCGAGGA AGTTCGTTCG AAAGCCCTTC TTCTTCAGGG GTTTTAGACG CAGGAGAGTA

+2    S  D  I  E   E  M  T   D  S  S   L  P  S  S   P  K  L   P  H  N   G  S  T  G   S  T  P
641   TCAGATATTG AGGAGATGAC GGATTCTTCT TTGCCTTCCT CACCAAAGTT ACCGCACAAT GGGTCCACAG GTTCCACCCC
      AGTCTATAAC TCCTCTACTG CCTAAGAAGA AACGGAAGGA GTGGTTTCAA TGGCGTGTTA CCCAGGTGTC CAAGGTGGGG

+2    L  L  R   N  S  H  S   N  S  L   I  S  E   C  M  D  S   E  A  E   T  V  M   Q  L  R
721   ACTGCTGAGG AATTCTCACT CCAACTCTCT AATTTCAGAA TGCATGGATA GTGAAGCTGA GACCGTCATG CAGCTCCGAA
      TGACGACTCC TTAAGAGTGA GGTTGAGAGA TTAAAGTCTT ACGTACCTAT CACTTCGACT CTGGCAGTAC GTCGAGGCTT

+2  N  E  L  R   D  K  E   M  K  L  T   D
801   ATGAGTTAAG AGACAAGGAG ATGAAGCTGA CGGATAT
      TACTCAATTC TCTGTTCCTC TACTTCGACT GCCTATA
```

11/24/97 02:54PM hh2UNC53 E2.3 14/3/97 F Page 1

```
  +1 R  S  P  S  D  A  G  R  S  S  G  D  E  S  K  K  P  L  P  S  S  S  R  T  P  T  A
   1 CGCTCCCCAT CAGATGCAGG CCGGAGCAGT GGTGACGAAT CCAAAAAGCC CCTCCCCAGC AGCTCTAGGA CACCTACTGC
     GCGAGGGGTA GTCTACGTCC GGCCTCGTCA CCACTGCTTA GGTTTTTCGG GGAGGGGTCG TCGAGATCCT GTGGATGACG
.......................................................................................
  +1  N  A  N  S  F  G  F  K  K  Q  S  G  S  A  A  G  L  A  M  I  T  A  S  G  V  T
  81 CAATGCCAAC AGCTTTGGGT TCAAGAAGCA GAGTGGTTCC GCCGCCGGCC TGGCCATGAT CACAGCCAGC GGGGTGACTG
     GTTACGGTTG TCGAAACCCA AGTTCTTCGT CTCACCAAGG CGGCGGCCGG ACCGGTACTA GTGTCGGTCG CCCCACTGAC
.......................................................................................
  +1 V  T  S  R  S  A  T  L  G  K  I  P  K  S  S  A  L  V  S  R  S  A  G  R  K  S  S
 161 TCACCAGCAG GTCAGCCACA CTGGGCAAAA TCCCAAAGTC ATCTGCACTC GTCAGTCGGT CTGCTGGTCG GAAGTCAAGT
     AGTGGTCGTC CAGTCGGTGT GACCCGTTTT AGGGTTTCAG TAGACGTGAG CAGTCAGCCA GACGACCAGC CTTCAGTTCA
.......................................................................................
  +1 M  D  G  A  Q  N  Q  D  D  G  Y  L  A  L  S  S  R  T  N  L  Q  Y  R  S  L  P  R
 241 ATGGATGGGG CTCAGAATCA GGATGACGGG TATCTAGCCC TAAGCTCCCG GACAAACCTT CAGTACCGGA GTTTGCCGAG
     TACCTACCCC GAGTCTTAGT CCTACTGCCC ATAGATCGGG ATTCGAGGGC CTGTTTGGAA GTCATGGCCT CAAACGGCTC
.......................................................................................
  +1  P  S  K  S  N  S  R  N  G  A  G  N  R  S  S  T  S  S  I  D  S  N  I  S  S  K
 321 GCCCAGTAAG TCCAACAGCC GGAACGGGGC TGGGAACAGG TCTAGCACCA GCAGCATAGA TTCCAACATT AGCAGCAAGT
     CGGGTCATTC AGGTTGTCGG CCTTGCCCCG ACCCTTGTCC AGATCGTGGT CGTCGTATCT AAGGTTGTAA TCGTCGTTCA
.......................................................................................
  +1 S  A  G  L  P  V  P  K  L  R  E  P  S  K  T  A  L  G  S  S  L  P  G  L  V  N  Q
 401 CCGCAGGCCT GCCAGTGCCC AAACTGAGGG AGCCTTCCAA AACAGCCCTA GGCAGCTCTC TACCAGGTCT GGTCAACCAA
     GGCGTCCGGA CGGTCACGGG TTTGACTCCC TCGGAAGGTT TTGTCGGGAT CCGTCGAGAG ATGGTCCAGA CCAGTTGGTT
.......................................................................................
  +1 T  D  K  E  K  G  I  S  S  D  N  E  S  V  A  S  C  N  S  V  K  V  N  P  A  A  Q
 481 ACAGATAAGG AGAAAGGCAT CTCATCAGAC AACGAGAGTG TGGCTTCCTG TAACTCGGTG AAAGTGAATC CGGCAGCCCA
     TGTCTATTCC TCTTTCCGTA GAGTAGTCTG TTGCTCTCAC ACCGAAGGAC ATTGAGCCAC TTTCACTTAG GCCGTCGGGT
.......................................................................................
  +1  P  V  S  S  P  A  Q  T  S  L  Q  P  G  A  K  Y  P  D  V  A  S  P  T  L  R  R
 561 GCCTGTGTCC AGTCCGGCTC AGACCAGTCT CCAGCCTGGA GCCAAGTACC CAGATGTGGC CTCTCCCACA CTCCGCAGAC
     CGGACACAGG TCAGGCCGAG TCTGGTCAGA GGTCGGACCT CGGTTCATGG GTCTACACCG GAGAGGGTGT GAGGCGTCTG
.......................................................................................
  +1 L  F  G  G  K  P  T  K  Q  V  P  I  A  T  A  E  N  M  K  N  S  V  V  I  S  N  P
 641 TCTTTGGTGG GAAGCCTACC AAGCAAGTGC CCATCGCCAC AGCTGAAAAC ATGAAAAATT CGGTGGTCAT CTCCAATCCT
     AGAAACCACC CTTCGGATGG TTCGTTCACG GGTAGCGGTG TCGACTTTTG TACTTTTTAA GCCACCAGTA GAGGTTAGGA
.......................................................................................
  +1 H  A  T  M  T  Q  Q  G  N  L  D  S  P  S  G  S  G  V  L  S  S  G  S  S  S  P  L
 721 CATGCCACCA TGACTCAGCA AGGTAACCTA GACTCCCCGT CAGGCAGTGG CGTCCTGAGC AGTGGGAGCA GCAGTCCTCT
     GTACGGTGGT ACTGAGTCGT TCCATTGGAT CTGAGGGGCA GTCCGTCACC GCAGGACTCG TCACCCTCGT CGTCAGGAGA
.......................................................................................
  +1  Y  S  K  N  V  D  L  N  Q  S  P  L  A  S  S  P  S  S  A  H  S  A  P  S  N  S
 801 CTACAGCAAG AATGTGGACC TCAACCAGTC TCCGCTAGCC TCCAGCCCCA GCTCAGCCCA CTCGGCCCCT TCCAACAGCC
     GATGTCGTTC TTACACCTGG AGTTGGTCAG AGGCGATCGG AGGTCGGGGT CGAGTCGGGT GAGCCGGGGA AGGTTGTCGG
.......................................................................................
  +1 L  T  W  G  T  N  A  S  S  S  S  A  V  S  K  D  G  L  G  F  Q  S  V  S  S  L  H
 881 TCACCTGGGG CACCAACGCC AGCAGCTCCT CCGCAGTTAG CAAGGATGGC CTGGGCTTTC AGTCTGTCAG CAGCCTCCAC
     AGTGGACCCC GTGGTTGCGG TCGTCGAGGA GGCGTCAATC GTTCCTACCG GACCCGAAAG TCAGACAGTC GTCGGAGGTG
.......................................................................................
  +1 T  S  C  E  S  I  D  I  S  L  S  S  G  G  V  P  S  H  N  S  S  T  G  L  I  A  S
 961 ACCAGCTGTG AGTCCATCGA CATCTCCCTC AGCAGTGGAG GGGTCCCCAG CCACAATTCT TCCACTGGCC TCATCGCCTC
     TGGTCGACAC TCAGGTAGCT GTAGAGGGAG TCGTCACCTC CCCAGGGGTC GGTGTTAAGA AGGTGACCGG AGTAGCGGAG
.......................................................................................
  +1  S  K  D  D  S  L  T  P  F  V  R  T  N  S  V  K  T  T  L  S  E  S  P  L  S  S
1041 CTCCAAGGAC GACTCCTTGA CTCCCTTTGT CAGAACTAAC AGTGTGAAGA CCACACTGTC AGAAAGCCCT CTCTCTTCCC
     GAGGTTCCTG CTGAGGAACT GAGGGAAACA GTCTTGATTG TCACACTTCT GGTGTGACAG TCTTTCGGGA GAGAGAAGGG
.......................................................................................
  +1 P  A  A  S  P  K  F  C  R  S  T  L  P  R  K  Q  D  S  D  P  H  L  D  R  N  T  L
1121 CTGCTGCTAG CCCTAAGTTC TGCAGAAGTA CTCTGCCCAG GAAACAGGAC AGTGACCCGC ACCTTGATAG GAACACTTTG
     GACGACGATC GGGATTCAAG ACGTCTTCAT GAGACGGGTC CTTTGTCCTG TCACTGGGCG TGGAACTATC CTTGTGAAAC
.......................................................................................
  +1  P  K  K  G  L  R  Y  T  P  T  S  Q  L  R  T  Q  E  D  A  K
1201 CCTAAGAAAG GACTCAGGTA TACTCCCACC TCCCAGCTTC GCACGCAAGA AGATGCAAAA
     GGATTCTTTC CTGAGTCCAT ATGAGGGTGG AGGGTCGAAG CGTGCGTTCT TCTACGTTTT
.......................................................................................
```

```
11/24/97 02:54PM            hh2UNC53 C2.3 14/3/97                Page 1

+1   R  S  P  S  D  A  G  R  S  S  G  D  E  S  L  K  P  L  P  S  S  S  R  T  P  T  A
1    CGCTCCCCAT CAGATGCAGG CCGGAGCAGT GGTGACGAAT CCTTAAAGCC CCTCCCCAGC AGCTCTAGGA CACCTACTGC
     GCGAGGGGTA GTCTACGTCC GGCCTCGTCA CCACTGCTTA GGAATTTCGG GGAGGGGTCG TCGAGATCCT GTGGATGACG
..........................................................................................
+1    N  A  N  S  F  G  F  K  K  Q  S  G  S  A  A  G  L  A  M  I  T  A  S  G  V  T
81   CAATGCCAAC AGCTTTGGGT TCAAGAAGCA GAGTGGTTCC GCCGCCGGCC TGGCCATGAT CACAGCCAGC GGGGTGACTG
     GTTACGGTTG TCGAAACCCA AGTTCTTCGT CTCACCAAGG CGGCGGCCGG ACCGGTACTA GTGTCGGTCG CCCCACTGAC
..........................................................................................
+1   V  T  S  R  S  A  T  L  G  K  I  P  K  S  S  A  L  V  S  R  S  A  G  R  K  S  S
161  TCACCAGCAG GTCAGCCACA CTGGGCAAAA TCCCAAAGTC ATCTGCACTC GTCAGTCGGT CTGCTGGTCG GAAGTCAAGT
     AGTGGTCGTC CAGTCGGTGT GACCCGTTTT AGGGTTTCAG TAGACGTGAG CAGTCAGCCA GACGACCAGC CTTCAGTTCA
..........................................................................................
+1   M  D  G  A  Q  N  Q  D  D  G  Y  L  A  L  S  S  R  T  N  L  Q  Y  R  S  L  P  R
241  ATGGATGGGG CTCAGAATCA GGATGACGGG TATCTAGCCC TAAGCTCCCG GACAAACCTT CAGTACCGGA GTTTGCCGAG
     TACCTACCCC GAGTCTTAGT CCTACTGCCC ATAGATCGGG ATTCGAGGGC CTGTTTGGAA GTCATGGCCT CAAACGGCTC
..........................................................................................
+1    P  S  K  S  N  S  R  N  G  A  G  N  R  S  S  T  S  S  I  D  S  N  I  S  S  K
321  GCCCAGTAAG TCCAACAGCC GGAACGGGGC TGGGAACAGG TCTAGCACCA GCAGCATAGA TTCCAACATT AGCAGCAAGT
     CGGGTCATTC AGGTTGTCGG CCTTGCCCCG ACCCTTGTCC AGATCGTGGT CGTCGTATCT AAGGTTGTAA TCGTCGTTCA
..........................................................................................
+1   S  A  G  L  P  V  P  K  L  R  E  P  S  K  T  A  L  G  S  S  L  P  G  L  V  N  Q
401  CCGCAGGCCT GCCAGTGCCC AAACTGAGGG AGCCTTCCAA AACAGCCCTA GGCAGCTCTC TACCAGGTCT GGTCAACCAA
     GGCGTCCGGA CGGTCACGGG TTTGACTCCC TCGGAAGGTT TTGTCGGGAT CCGTCGAGAG ATGGTCCAGA CCAGTTGGTT
..........................................................................................
+1   T  D  K  E  K  G  I  S  S  D  N  E  S  V  A  S  C  N  S  V  K  V  N  P  A  A  Q
481  ACAGACAAGG AGAAAGGCAT CTCATCAGAC AACGAGAGTG TGGCTTCCTG TAACTCGGTG AAAGTGAATC CGGCAGCCCA
     TGTCTGTTCC TCTTTCCGTA GAGTAGTCTG TTGCTCTCAC ACCGAAGGAC ATTGAGCCAC TTTCACTTAG GCCGTCGGGT
..........................................................................................
+1    P  V  S  S  P  A  Q  T  S  L  Q  P  G  A  K  Y  P  D  V  A  S  P  T  L  R  R
561  GCCTGTGTCC AGTCCGGCTC AGACCAGTCT CCAGCCTGGA GCCAAGTACC CAGATGTGGC CTCTCCCACA CTCCGCAGAC
     CGGACACAGG TCAGGCCGAG TCTGGTCAGA GGTCGGACCT CGGTTCATGG GTCTACACCG GAGAGGGTGT GAGGCGTCTG
..........................................................................................
+1   L  F  G  G  K  P  T  K  Q  V  P  I  A  T  A  E  N  M  K  N  S  V  V  I  S  N  P
641  TCTTTGGTGG GAAGCCTACC AAGCAAGTGC CCATCGCCAC AGCTGAAAAC ATGAAAAATT CGGTGGTCAT CTCCAATCCT
     AGAAACCACC CTTCGGATGG TTCGTTCACG GGTAGCGGTG TCGACTTTTG TACTTTTTAA GCCACCAGTA GAGGTTAGGA
..........................................................................................
+1   H  A  T  M  T  Q  Q  G  N  L  D  S  P  S  G  S  G  V  L  S  S  G  S  S  S  P  L
721  CATGCCACCA TGACTCAGCA AGGTAACCTA GACTCCCCGT CAGGCAGTGG CGTCCTGAGC AGTGGGAGCA GCAGTCCTCT
     GTACGGTGGT ACTGAGTCGT TCCATTGGAT CTGAGGGGCA GTCCGTCACC GCAGGACTCG TCACCCTCGT CGTCAGGAGA
..........................................................................................
+1    Y  S  K  N  V  D  L  N  Q  S  P  L  A  S  S  P  S  S  A  H  S  A  P  S  N  S
801  CTACAGCAAG AATGTGGACC TCAACCAGTC TCCGCTAGCC TCCAGCCCCA GCTCAGCCCA CTCGGCCCCT TCCAACAGCC
     GATGTCGTTC TTACACCTGG AGTTGGTCAG AGGCGATCGG AGGTCGGGGT CGAGTCGGGT GAGCCGGGGA AGGTTGTCGG
..........................................................................................
+1   L  T  W  G  T  N  A  S  S  S  S  A  V  S  K  D  G  L  G  F  Q  S  V  S  S  L  H
881  TCACCTGGGG CACCAACGCC AGCAGCTCCT CCGCAGTTAG CAAGGATGGC CTGGGCTTTC AGTCTGTCAG CAGCCTCCAC
     AGTGGACCCC GTGGTTGCGG TCGTCGAGGA GGCGTCAATC GTTCCTACCG GACCCGAAAG TCAGACAGTC GTCGGAGGTG
..........................................................................................
+1   T  S  C  E  S  I  D  I  S  L  S  S  G  G  V  P  S  H  N  S  S  T  G  L  I  A  S
961  ACCAGCTGTG AGTCCATCGA CATCTCCCTC AGCAGTGGAG GGGTCCCCAG CCACAATTCT TCCACTGGCC TCATCGCCTC
     TGGTCGACAC TCAGGTAGCT GTAGAGGGAG TCGTCACCTC CCCAGGGGTC GGTGTTAAGA AGGTGACCGG AGTAGCGGAG
..........................................................................................
+1    S  K  D  D  S  L  T  P  F  V  R  T  N  S  V  K  T  T  L  S  E  R  Y  T  P  T
1041 CTCCAAGGAC GACTCCTTGA CTCCCTTTGT CAGAACTAAC AGTGTGAAGA CCACACTGTC AGAAAGGTAT ACTCCCACCT
     GAGGTTCCTG CTGAGGAACT GAGGGAAACA GTCTTGATTG TCACACTTCT GGTGTGACAG TCTTTCCATA TGAGGGTGGA
..........................................................................................
+1   S  Q  L  R  T  Q  E  D  A  K
1121 CCCAGCTTCG CACGCAAGAA GATGCAAAA
     GGGTCGAAGC GTGCGTTCTT CTACGTTTT
..........................................................................................
```

```
11/24/97 03:29PM                         E1.3-3                          Page 1

 +2                                                 S  W  D   D  S  S   S  V  S   S  G  I
  1  GGATCCACTA GTAACGGCCG CCAGTGTGCT GGAATTCGGC TTCAGCTGGG ACGACAGCAG CTCCGTCAGC AGCGGCATCA
     CCTAGGTGAT CATTGCCGGC GGTCACACGA CCTTAAGCCG AAGTCGACCC TGCTGTCGTC GAGGCAGTCG TCGCCGTAGT
...........................................................................................
 +2 S  D  T  I   D  N  L   S  T  D  D   I  N  T   S  S  S   I  S  S  Y   A  N  T   P  A  S
 81  GCGACACCAT AGACAACCTC AGCACTGATG ACATCAACAC CAGCTCCTCC ATCAGCTCTT ATGCCAACAC ACCTGCCTCC
     CGCTGTGGTA TCTGTTGGAG TCGTGACTAC TGTAGTTGTG GTCGAGGAGG TAGTCGAGAA TACGGTTGTG TGGACGGAGG
...........................................................................................
 +2  S  R  K  N   L  D  V   Q  T  D   A  E  K  H   S  Q  V   E  R  N   S  L  W  S   G  D
161  TCTCGAAAAA ACCTGGATGT GCAGACTGAT GCTGAGAAGC ACTCACAGGT GGAGAGGAAT TCCCTGTGGT CTGGTGATGA
     AGAGCTTTTT TGGACCTACA CGTCTGACTA CGACTCTTCG TGAGTGTCCA CCTCTCCTTA AGGGACACCA GACCACTACT
...........................................................................................
 +2   V  K  K   S  D  G  G   S  D  S   G  I  K   M  E  P  G   S  K  W   R  R  N   P  S  D
241  TGTCAAGAAA TCAGACGGAG GCTCAGACAG CGGCATAAAA ATGGAGCCAG GTTCCAAGTG GAGGCGGAAT CCTTCTGATG
     ACAGTTCTTT AGTCTGCCTC CGAGTCTGTC GCCGTATTTT TACCTCGGTC CAAGGTTCAC CTCCGCCTTA GGAAGACTAC
...........................................................................................
 +2 V  S  D  E   S  D  K   S  T  S  G   K  K  N   P  V  I   S  Q  T  G   S  W  R   R  G  M
321  TGTCTGACGA GTCCGACAAA AGCACGTCGG GCAAGAAGAA TCCTGTCATC TCCCAGACAG GCTCATGGCG GCGAGGCATG
     ACAGACTGCT CAGGCTGTTT TCGTGCAGCC CGTTCTTCTT AGGACAGTAG AGGGTCTGTC CGAGTACCGC CGCTCCGTAC
...........................................................................................
 +2  T  A  Q  V   G  I  T   M  P  R   T  K  P  S   A  P  A   G  A  L   K  T  P  G   T  G
401  ACAGCTCAGG TGGGCATCAC CATGCCAAGG ACGAAGCCTT CAGCCCCGGC AGGCGCACTG AAGACCCCAG GAACTGGAAA
     TGTCGAGTCC ACCCGTAGTG GTACGGTTCC TGCTTCGGAA GTCGGGGCCG TCCGCGTGAC TTCTGGGGTC CTTGACCTTT
...........................................................................................
 +2   T  D  D   A  K  V  S   E  K  G   R  L  S   P  K  A  S   Q  V  K   R  S  P
481  AACAGACGAC GCAAAGGTGT CTGAGAAAGG AAGGCTTTCT CCTAAAGCCT CCCAGGTGAA GCGCTCCCCA TC
     TTGTCTGCTG CGTTTCCACA GACTCTTTCC TTCCGAAAGA GGATTTCGGA GGGTCCACTT CGCGAGGGGT AG
...........................................................................................
```

```
11/24/97 03:34PM                 hh25'raceB2-1_CON                I            Page 1

 +1 S  W  R  R   G  M  T   A  Q  V   G  I  T  M   P  R  T   K  A  S   A  P  A  G   A  L
   1 TCATGGCGGC GAGGCATGAC AGCTCAGGTG GGCATCACCA TGCCAAGGAC GAAGGCTTCA GCCCCGGCAG GCGCACTGAA
     AGTACCGCCG CTCCGTACTG TCGAGTCCAC CCGTAGTGGT ACGGTTCCTG CTTCCGAAGT CGGGGCCGTC CGCGTGACTT
.........................................................................................
 +1   T  P  G   T  G  K  T   D  D  A   K  V  S   E  K  G  R   L  S  P   K  A  S   Q  V  K
  81 GACCCCAGGA ACTGGAAAAA CAGACGACGC AAAGGTGTCT GAGAAAGGAA GGCTTTCTCC TAAAGCCTCC CAGGTGAAGC
     CTGGGGTCCT TGACCTTTTT GTCTGCTGCG TTTCCACAGA CTCTTTCCTT CCGAAAGAGG ATTTCGGAGG GTCCACTTCG
.........................................................................................
 +1 R  S  P  S   D  A  G   R  S  S  G   D  E  S   K  K  P   L  P  S  S   S  R  T   P  T  A
 161 GCTCCCCATC AGATGCAGGC CGGAGCAGTG GTGACGAATC CAAAAAGCCC CTCCCCAGCA GCTCTAGGAC ACCTACTGCC
     CGAGGGGTAG TCTACGTCCG GCCTCGTCAC CACTGCTTAG GTTTTTCGGG GAGGGGTCGT CGAGATCCTG TGGATGACGG
.........................................................................................
 +1 N  A  N  S   F  G  F   K  K  Q   S  G  S  A   A  G  L   A  M  I   T  A  S  G   V  T
 241 AATGCCAACA GCTTTGGGTT CAAGAAGCAG AGTGGTTCCG CCGCCGGCCT GGCCATGATC ACAGCCAGCG GGGTGACTGT
     TTACGGTTGT CGAAACCCAA GTTCTTCGTC TCACCAAGGC GGCGGCCGGA CCGGTACTAG TGTCGGTCGC CCCACTGACA
.........................................................................................
 +1   T  S  R   S  A  T  L   G  K  I   P  K  S   S  A  L  V   S  R  S   A  G  R   K  S  S
 321 CACCAGCAGG TCAGCCACAC TGGGCAAAAT CCCAAAGTCA TCTGCACTCG TCAGTCGGTC TGCTGGTCGG AAGTCAAGTA
     GTGGTCGTCC AGTCGGTGTG ACCCGTTTTA GGGTTTCAGT AGACGTGAGC AGTCAGCCAG ACGACCAGCC TTCAGTTCAT
.........................................................................................
 +1 M  D  G  A   Q  N  Q   D  D  G  Y   L  A  L   S  S  R   T  N  L  Q   Y  R  S   L  P  R
 401 TGGATGGGGC TCAGAATCAG GATGACGGGT ATCTAGCCCT AAGCTCCCGG ACAAACCTTC AGTACCGGAG TTTGCCGAGG
     ACCTACCCCG AGTCTTAGTC CTACTGCCCA TAGATCGGGA TTCGAGGGCC TGTTTGGAAG TCATGGCCTC AAACGGCTCC
.........................................................................................
 +1 P  S  K  S   N  S  R   N  G  A   G  N  R  S   S  T  S   S  I  D   S  N  I  S   S  K
 481 CCCAGTAAGT CCAACAGCCG GAACGGGGCT GGGAACAGGT CTAGCACCAG CAGCATAGAT TCCAACATTA GCAGCAAGTC
     GGGTCATTCA GGTTGTCGGC CTTGCCCCGA CCCTTGTCCA GATCGTGGTC GTCGTATCTA AGGTTGTAAT CGTCGTTCAG
.........................................................................................
 +1   A  G  L   P  V  P  K   L  R  E   P  S  K   T  A  L  G   S  S  L   P  G  L   V  N  Q
 561 CGCAGGCCTG CCAGTGCCCA AACTGAGGGA GCCTTCCAAA ACAGCCCTAG GCAGCTCTCT ACCAGGTCTG GTCAACCAAA
     GCGTCCGGAC GGTCACGGGT TTGACTCCCT CGGAAGGTTT TGTCGGGATC CGTCGAGAGA TGGTCCAGAC CAGTTGGTTT
.........................................................................................
 +1 T  D  K  E   K  G  I   S  S  D  N   E  S  V   A  S  C   N  S  V  K   V  N  P   A  A  Q
 641 CAGATAAGGA GAAAGGCATC TCATCAGACA ACGAGAGTGT GGCTTCCTGT AACTCGGTGA AAGTGAATCC GGCAGCCCAG
     GTCTATTCCT CTTTCCGTAG AGTAGTCTGT TGCTCTCACA CCGAAGGACA TTGAGCCACT TTCACTTAGG CCGTCGGGTC
.........................................................................................
 +1 P  V  S  S   P  A  Q   T  S  L   Q  P  G  A   K  Y  P   D  V  A   S  P  T  L   R  R
 721 CCTGTGTCCA GTCCGGCTCA GACCAGTCTC CAGCCTGGAG CCAAGTACCC AGATGTGGCC TCTCCCACAC TCCGCAGACT
     GGACACAGGT CAGGCCGAGT CTGGTCAGAG GTCGGACCTC GGTTCATGGG TCTACACCGG AGAGGGTGTG AGGCGTCTGA
.........................................................................................
 +1   F  G  G   K  P  T  K   Q  V  P   I  A  T   A  E  N  M   K  N  S   V  V  I   S  N  P
 801 CTTTGGTGGG AAGCCTACCA AGCAAGTGCC CATCGCCACA GCTGAAAACA TGAAAAATTC GGTGGTCATC TCCAATCCTC
     GAAACCACCC TTCGGATGGT TCGTTCACGG GTAGCGGTGT CGACTTTTGT ACTTTTTAAG CCACCAGTAG AGGTTAGGAG
.........................................................................................
 +1 H  A  T  M   T  Q  Q   G  N  L  D   S  P  S   G  S  G   V  L  S  S   G  S  S   S  P  L
 881 ATGCCACCAT GACTCAGCAA GGTAACCTAG ACTCCCCGTC AGGCAGTGGC GTCCTGAGCA GTGGGAGCAG CAGTCCTCTC
     TACGGTGGTA CTGAGTCGTT CCATTGGATC TGAGGGGCAG TCCGTCACCG CAGGACTCGT CACCCTCGTC GTCAGGAGAG
.........................................................................................
 +1 Y  S  K  N   V  D  L   N  Q  S   P  L  A  S   S  P  S   S  A  H   S  A  P  S   N  S
 961 TACAGCAAGA ATGTGGACCT CAACCAGTCT CCGCTAGCCT CCAGCCCCAG CTCAGCCCAC TCGGCCCCTT CCAACAGCCT
     ATGTCGTTCT TACACCTGGA GTTGGTCAGA GGCGATCGGA GGTCGGGGTC GAGTCGGGTG AGCCGGGGAA GGTTGTCGGA
.........................................................................................
 +1   T  W  G   T  N  A  S   S  S  S   A  V  S   K  D  G  L   G  F  Q   S  V  S   S  L  H
1041 CACCTGGGGC ACCAACGCCA GCAGCTCCTC CGCAGTTAGC AAGGATGGCC TGGGCTTTCA GTCTGTCAGC AGCCTCCACA
     GTGGACCCCG TGGTTGCGGT CGTCGAGGAG GCGTCAATCG TTCCTACCGG ACCCGAAAGT CAGACAGTCG TCGGAGGTGT
.........................................................................................
 +1 T  S  C  E   S  I  D   I  S  L  S   S  G  G   V  P  S   H  N  S  S   T  G  L   I  A  S
1121 CCAGCTGTGA GTCCATCGAC ATCTCCCTCA GCAGTGGAGG GGTCCCCAGC CACAATTCTT CCACTGGCCT CATCGCCTCC
     GGTCGACACT CAGGTAGCTG TAGAGGGAGT CGTCACCTCC CCAGGGGTCG GTGTTAAGAA GGTGACCGGA GTAGCGGAGG
.........................................................................................
 +1 S  K  D  D   S  L  T   P  F  V   R  T  N  S   V  K  T   T  L  S   E  S  D  P   H  L
1201 TCCAAGGACG ACTCCTTGAC TCCCTTTGTC AGAACTAACA GTGTGAAGAC CACACTGTCA GAAAGTGACC CGCACCTTGA
     AGGTTCCTGC TGAGGAACTG AGGGAAACAG TCTTGATTGT CACACTTCTG GTGTGACAGT CTTTCACTGG GCGTGGAACT
.........................................................................................
```

```
11/24/97 03:34PM                 hh25'raceB2-1_CON                             Page 2

 +1   R  N  T   L  P  K  K   G  L  R   V
1281 TAGGAACACT TTGCCTAAGA AAGGACTCAG GGTAA
     ATCCTTGTGA AACGGATTCT TTCCTGAGTC CCATT
.........................................................................................
```

```
+2   P V I  S Q T G  S W R  R G M  T A Q V  G I T  M P R  T K A
1    TCCTGTCATC TCCCAGACAG GCTCATGGCG GCGAGGCATG ACAGCTCAGG TGGGCATCAC CATGCCAAGG ACGAAGGCTT
     AGGACAGTAG AGGGTCTGTC CGAGTACCGC CGCTCCGTAC TGTCGAGTCC ACCCGTAGTG GTACGGTTCC TGCTTCCGAA
.......................................................................................
+2 S A P A  G A L  K T P G  T G K  T D D  A K V S  E K G  R L S
81   CAGCCCCGGC AGGCGCACTG AAGACCCCAG GAACTGGAAA AACAGACGAC GCAAAGGTGT CTGAGAAAGG AAGGCTTTCT
     GTCGGGGCCG TCCGCGTGAC TTCTGGGGTC CTTGACCTTT TTGTCTGCTG CGTTTCCACA GACTCTTTCC TTCCGAAAGA
.......................................................................................
+2 P K A S  Q V K  R S P  S D A G  R S S  G D E  S K K P  L P
161  CCTAAAGCCT CCCAGGTGAA GCGCTCCCCA TCAGATGCAG GCCGGAGCAG TGGTGACGAA TCCAAAAAGC CCCTCCCCAG
     GGATTTCGGA GGGTCCACTT CGCGAGGGGT AGTCTACGTC CGGCCTCGTC ACCACTGCTT AGGTTTTTCG GGGAGGGGTC
.......................................................................................
+2  S S R  T P T A  N A N  S F G  F K K Q  S G S  A A G  L A M
241  CAGCTCTAGG ACACCTACTG CCAATGCCAA CAGCTTTGGG TTCAAGAAGC AGAGTGGTTC CGCCGCCGGC CTGGCCATGA
     GTCGAGATCC TGTGGATGAC GGTTACGGTT GTCGAAACCC AAGTTCTTCG TCTCACCAAG GCGGCGGCCG GACCGGTACT
.......................................................................................
+2 I T A S  G V T  V T S R  S A T  L G K  I P K S  S A L  V S R
321  TCACAGCCAG CGGGGTGACT GTCACCAGCA GGTCAGCCAC ACTGGGCAAA ATCCCAAAGT CATCTGCACT CGTCAGTCGG
     AGTGTCGGTC GCCCCACTGA CAGTGGTCGT CCAGTCGGTG TGACCCGTTT TAGGGTTTCA GTAGACGTGA GCAGTCAGCC
.......................................................................................
+2 S A G R  K S S  M D G  A Q N Q  D D G  Y L A  L S S R  T N
401  TCTGCTGGTC GGAAGTCAAG TATGGATGGG GCTCAGAATC AGGATGACGG GTATCTAGCC CTAAGCTCCC GGACAAACCT
     AGACGACCAG CCTTCAGTTC ATACCTACCC CGAGTCTTAG TCCTACTGCC CATAGATCGG GATTCGAGGG CCTGTTTGGA
.......................................................................................
+2  Q Y R  S L P R  P S K  S N S  R N G A  G N R  S S T  S S I
481  TCAGTACCGG AGTTTGCCGA GGCCCAGTAA GTCCAACAGC CGGAACGGGG CTGGGAACAG GTCTAGCACC AGCAGCATAG
     AGTCATGGCC TCAAACGGCT CCGGGTCATT CAGGTTGTCG GCCTTGCCCC GACCCTTGTC CAGATCGTGG TCGTCGTATC
.......................................................................................
+2 D S N I  S S K  S A G L  P V P  K L R  E P S K  T A L  G S S
561  ATTCCAACAT TAGCAGCAAG TCCGCAGGCC TGCCAGTGCC CAAACTGAGG GAGCCTTCCA AAACAGCCCT AGGCAGCTCT
     TAAGGTTGTA ATCGTCGTTC AGGCGTCCGG ACGGTCACGG GTTTGACTCC CTCGGAAGGT TTTGTCGGGA TCCGTCGAGA
.......................................................................................
+2  L P G L  V N Q  T D K  E K G I  S S D  N E S  V A S C  N S
641  CTACCAGGTC TGGTCAACCA AACAGATAAG GAGAAAGGCA TCTCATCAGA CAACGAGAGT GTGGCTTCCT GTAACTCGGT
     GATGGTCCAG ACCAGTTGGT TTGTCTATTC CTCTTTCCGT AGAGTAGTCT GTTGCTCTCA CACCGAAGGA CATTGAGCCA
.......................................................................................
+2  K V N  P A A Q  P V S  S P A  Q T S L  Q P G  A K Y  P D V
721  GAAAGTGAAT CCGGCAGCCC AGCCTGTGTC CAGTCCGGCT CAGACCAGTC TCCAGCCTGG AGCCAAGTAC CCAGATGTGG
     CTTTCACTTA GGCCGTCGGG TCGGACACAG GTCAGGCCGA GTCTGGTCAG AGGTCGGACC TCGGTTCATG GGTCTACACC
.......................................................................................
+2 A S P T  L R R  L F G G  K P T  K Q V  P I A T  A E N  M K N
801  CCTCTCCCAC ACTCCGCAGA CTCTTTGGTG GGAAGCCTAC CAAGCAAGTG CCCATCGCCA CAGCTGAAAA CATGAAAAAT
     GGAGAGGGTG TGAGGCGTCT GAGAAACCAC CCTTCGGATG GTTCGTTCAC GGGTAGCGGT GTCGACTTTT GTACTTTTTA
.......................................................................................
+2  S V V I  S N P  H A T  H T Q Q  G H L  D S P  S G S G  V L
881  TCGGTGGTCA TCTCCAATCC TCATGCCACC ATGACTCAGC AAGGTAACCT AGACTCCCCG TCAGGCAGTG GCGTCCTGAG
     AGCCACCAGT AGAGGTTAGG AGTACGGTGG TACTGAGTCG TTCCATTGGA TCTGAGGGGC AGTCCGTCAC CGCAGGACTC
.......................................................................................
+2  S G S  S S P L  Y S K  N V D  L N Q S  P L A  S S P  S S A
961  CAGTGGGAGC AGCAGTCCTC TCTACAGCAA GAATGTGGAC CTCAACCAGT CTCCGCTAGC CTCCAGCCCC AGCTCAGCCC
     GTCACCCTCG TCGTCAGGAG AGATGTCGTT CTTACACCTG GAGTTGGTCA GAGGCGATCG GAGGTCGGGG TCGAGTCGGG
.......................................................................................
+2 H S A P  S N S  L T W G  T N A  S S S  S A V S  K D G  L G F
1041 ACTCGGCCCC TTCCAACAGC CTCACCTGGG GCACCAACGC CAGCAGCTCC TCCGCAGTTA GCAAGGATGG CCTGGGCTTT
     TGAGCCGGGG AAGGTTGTCG GAGTGGACCC CGTGGTTGCG GTCGTCGAGG AGGCGTCAAT CGTTCCTACC GGACCCGAAA
.......................................................................................
+2  Q S V S  S L H  T S C  X S I D  I S L  X S G  G V P X  H N
1121 CAGTCTGTCA GCAGCCTCCA CACCAGCTGT GADTCCATCG ACATCTCCCT CARCAGTGGA GGGGTCCCCA RCCACAATTC
     GTCAGACAGT CGTCGGAGGT GTGGTCGACA CTHAGGTAGC TGTAGAGGGA GTYGTCACCT CCCCAGGGGT YGGTGTTAAG
.......................................................................................
+2  S T G  L I A S  S K D  X S L  T P F V  R T N  S V K  T T L
1201 TTCCACTGGC CTCATCGCCT CCTCCAAGGA CRACTCCTTG ACTCCCTTTG TCAGAACTAA CAGTGTGAAG ACCACACTGT
     AAGGTGACCG GAGTAGCGGA GGAGGTTCCT GYTGAGGAAC TGAGGGAAAC AGTCTTGATT GTCACACTTC TGGTGTGACA
.......................................................................................
```



```
+2 S E S P  L S S  P A A S  P K F  C R X  X L P R  K Q D  S D P
1281 CAGAAAGCCC TCTCTCTTCC CCTGCTGCTA GCCCTAAGTT CTGCAGAANT WCTCTGCCCA GGAAACAGGA CAGTGACCCG
     GTCTTTCGGG AGAGAGAAGG GGACGACGAT CGGGATTCAA GACGTCTTNA WGAGACGGGT CCTTTGTCCT GTCACTGGGC
.......................................................................................
+2  H L D R  N T L  P K K  G L R X
1361 CACCTTGATA GGAACACTTT GCCTAAGAAA GGACTCAGGT WT
     GTGGAACTAT CCTTGTGAAA CGGATTCTTT CCTGAGTCCA WA
.......................................................................................
```

```
+2    L  R  Q   N  L  E  E   T  M  S   S  L  R   G  T  Q  V   T  H  S   T  L  E   T  T  F
   1  TCTGCGGCAG AATTTGGAGG AAACCATGTC CAGTTTAAGG GGAACTCAGG TTACACACAG CACATTGGAA ACCACGTTTG
      AGACGCCGTC TTAAACCTCC TTTGGTACAG GTCAAATTCC CCTTGAGTCC AATGTGTGTC GTGTAACCTT TGGTGCAAAC

+2 D  T  N  V   T  T  E   M  S  G  R   S  I  L   S  L  T   G  R  P  T   P  L  S   W  R  L
  81  ACACCAATGT CACCACGGAG ATGAGTGGCC GTAGCATACT CAGCTTGACA GGGAGGCCCA CACCTCTGTC CTGGAGACTG
      TGTGGTTACA GTGGTGCCTC TACTCACCGG CATCGTATGA GTCGAACTGT CCCTCCGGGT GTGGAGACAG GACCTCTGAC

+2  G  Q  S  S   P  R  L   Q  A  G   D  A  P  S   M  G  N   G  Y  P   P  R  A  N   A  S
 161  GGCCAGTCCA GCCCTCGGCT CCAAGCAGGA GACGCCCCCT CAATGGGCAA TGGGTATCCC CCTCGAGCCA ACGCCAGCAG
      CCGGTCAGGT CGGGAGCCGA GGTTCGTCCT CTGCGGGGGA GTTACCCGTT ACCCATAGGG GGAGCTCGGT TGCGGTCGTC

+2   F  I  N   T  E  S  G   R  Y  V   Y  S  A   P  L  R  R   Q  L  A   S  R  G   S  S  V
 241  GTTCATCAAC ACTGAGTCAG GTCGCTATGT GTACTCCGCC CCTCTGAGAA GGCAGCTGGC CTCCCGGGGC AGTAGTGTCT
      CAAGTAGTTG TGACTCAGTC CAGCGATACA CATGAGGCGG GGAGACTCTT CCGTCGACCG GAGGGCCCCG TCATCACAGA

+2 C  H  V  D   V  S  D   K  A  G  D   E  M  D   L  E  G   I  S  M  D   A  P  G   Y  M  S
 321  GCCATGTGGA CGTCTCAGAC AAGGCAGGAG ATGAGATGGA CCTGGAAGGC ATCAGCATGG ACGCCCCCGG CTACATGAGC
      CGGTACACCT GCAGAGTCTG TTCCGTCCTC TACTCTACCT GGACCTTCCG TAGTCGTACC TGCGGGGGCC GATGTACTCG

+2  D  G  D  V   L  S  K   N  I  R   T  D  D  I   T  S  G   Y  M  T   D  G  G  L   G  L
 401  GATGGGGATG TTCTGAGCAA GAACATCCGG ACCGATGACA TTACAAGCGG ATACATGACT GATGGTGGAC TTGGCCTCTA
      CTACCCCTAC AAGACTCGTT CTTGTAGGCC TGGCTACTGT AATGTTCGCC TATGTACTGA CTACCACCTG AACCGGAGAT

+2   T  R  R   L  N  R  L   P  D  G   M  A  V   V  R  E  T   L  Q  R   N  T  S   L  G  L
 481  TACCCGTCGC CTGAACCGGC TCCCTGATGG GATGGCTGTG GTACGGGAGA CCCTGCAACG AAATACCTCC CTGGGCCTCG
      ATGGGCAGCG GACTTGGCCG AGGGACTACC CTACCGACAC CATGCCCTCT GGGACGTTGC TTTATGGAGG GACCCGGAGC

+2 G  D  A  D   S  W  D   D  S  S  S   V  S  S   G  I  S   D  T  I  D   N  L  S   T  D  D
 561  GAGACGCTGA CAGCTGGGAC GACAGCAGCT CCGTCAGCAG CGGCATCAGC GACACCATAG ACAACCTCAG CACTGATGAC
      CTCTGCGACT GTCGACCCTG CTGTCGTCGA GGCAGTCGTC GCCGTAGTCG CTGTGGTATC TGTTGGAGTC GTGACTACTG

+2  I  N  T  S   S  S  I   S  S  Y   A  N  T  P   A  S  S   R  K  N   L  D  V  Q   T  D
 641  ATCAACACCA GCTCCTCCAT CAGCTCTTAT GCCAACACAC CTGCCTCCTC TCGAAAAAAC CTGGATGTGC AGACTGATGC
      TAGTTGTGGT CGAGGAGGTA GTCGAGAATA CGGTTGTGTG GACGGAGGAG AGCTTTTTTG GACCTACACG TCTGACTACG

+2   E  K  H   S  Q  V  E   R  N  S   L  W  S   G  D  D  V   K  K  S   D  G  G   S  D  S
 721  TGAGAAGCAC TCACAGGTGG AGAGGAATTC CCTGTGGTCT GGTGATGATG TCAAGAAATC AGACGGAGGC TCAGACAGCG
      ACTCTTCGTG AGTGTCCACC TCTCCTTAAG GGACACCAGA CCACTACTAC AGTTCTTTAG TCTGCCTCCG AGTCTGTCGC

+2 G  I  K  M   E  P  G   S  K  W  R   R  N  P   S  D  V   S  D  E  S   D  K  S   T  S  G
 801  GCATAAAAAT GGAGCCAGGT TCCAAGTGGA GGCGGAATCC TTCTGATGTG TCTGACGAGT CCGACAAAAG CACGTCGGGC
      CGTATTTTTA CCTCGGTCCA AGGTTCACCT CCGCCTTAGG AAGACTACAC AGACTGCTCA GGCTGTTTTC GTGCAGCCCG

+2   K  K  N   P  V  I  S   Q  T  G   S  W  R  R   G  M  T   A  Q  V   G  I  T  M   P  S T
 881  AAGAAGAATC CTGTCATCTC CCAGACAGGC TCATGGCGGC GAGGCATGAC AGCTCAGGTG GGCATCACCA TGCCAAGTAC
      TTCTTCTTAG GACAGTAGAG GGTCTGTCCG AGTACCGCCG CTCCGTACTG TCGAGTCCAC CCGTAGTGGT ACGGTTCATG

+2
+1     D  V  A   S  P  T   L  R  R   L  F  G  G   K  P  T   K  Q  V   P  I  A  T   A  E
 961  CCAGATGTGG CCTCTCCCAC ACTCCGCAGA CTCTTTGGTG GGAAGCCTAC CAAGCAAGTG CCCATCGCCA CAGCTGAAAA
      GGTCTACACC GGAGAGGGTG TGAGGCGTCT GAGAAACCAC CCTTCGGATG GTTCGTTCAC GGGTAGCGGT GTCGACTTTT

+1   M  K  N   S  V  V  I   S  N  P   H  A  T   M  T  Q  Q   G  N  L   D  S  P   S  G  S
1041  CATGAAAAAT TCGGTGGTCA TCTCCAATCC TCATGCCACC ATGACTCAGC AAGGTAACCT AGACTCCCCG TCAGGCAGTG
      GTACTTTTTA AGCCACCAGT AGAGGTTAGG AGTACGGTGG TACTGAGTCG TTCCATTGGA TCTGAGGGGC AGTCCGTCAC

+1 G  V  L  S   S  G  S   S  S  P  L   Y  S  K   N  V  D   L  N  Q  S   P  L  A   S  S  P
1121  GCGTCCTGAG CAGTGGGAGC AGCAGTCCTC TCTACAGCAA GAATGTGGAC CTCAACCAGT CTCCGCTAGC CTCCAGCCCC
      CGCAGGACTC GTCACCCTCG TCGTCAGGAG AGATGTCGTT CTTACACCTG GAGTTGGTCA GAGGCGATCG GAGGTCGGGG

+1  S  S  A  H   S  A  P   S  N  S   L  T  W  G   T  N  A   S  S  S   S  A  V  S   K  D
1201  AGCTCAGCCC ACTCGGCCCC TTCCAACAGC CTCACCTGGG GCACCAACGC CAGCAGCTCC TCCGCAGTTA GCAAGGATGG
      TCGAGTCGGG TGAGCCGGGG AAGGTTGTCG GAGTGGACCC CGTGGTTGCG GTCGTCGAGG AGGCGTCAAT CGTTCCTACC
```



```
+1    L  G  F   Q  S  V  S   S  L  H   T  S  C   E  S  I  D   I  S  L   S  S  G   G  V  P
1281  CCTGGGCTTT CAGTCTGTCA GCAGCCTCCA CACCAGCTGT GAGTCCATCG ACATCTCCCT CAGCAGTGGA GGGGTCCCCA
      GGACCCGAAA GTCAGACAGT CGTCGGAGGT GTGGTCGACA CTCAGGTAGC TGTAGAGGGA GTCGTCACCT CCCCAGGGGT

+1 S  H  N  S   S  T  G   L  I  A  S   S  K  D   D  S  L   T  P  F  V   R  T  N   S  V  K
1361  GCCACAATTC TTCCACTGGC CTCATCGCCT CCTCCAAGGA CGACTCCTTG ACTCCCTTTG TCAGAACTAA CAGTGTGAAG
      CGGTGTTAAG AAGGTGACCG GAGTAGCGGA GGAGGTTCCT GCTGAGGAAC TGAGGGAAAC AGTCTTGATT GTCACACTTC

+1  T  T  L  S   E  S  D   P  H  L   D  R  N  T   L  P  K   K  G  S   G
1441  ACCACACTGT CAGAAAGTGA CCCGCACCTT GATAGGAACA CTTTGCCTAA GAAAGGATCA GGTAA
      TGGTGTGACA GTCTTTCACT GGGCGTGGAA CTATCCTTGT GAAACGGATT CTTTCCTAGT CCATT
```

```
+2                                                  A R A G  E S G  S S S  T P T
         EcoRI
         -------
   1  CTGCAGAATT CGGCTTACTC ACTATAGGGC TCGAGCGGCC GCCCGGGCAG GTGAGAGCGG CAGCAGCTCC ACCCCTACTA
      GACGTCTTAA GCCGAATGAG TGATATCCCG AGCTCGCCGG CGGGCCCGTC CACTCTCGCC GTCGTCGAGG TGGGGATGAT
.............................................................................................
+2 N  C  S  T   S  S  A   I  P  Q  P   G  A  A   T  K  P   W  R  S  K   S  L  S   V  K  H
  81  ATTGCAGTAC CTCCTCGGCC ATCCCGCAGC CCGGTGCAGC CACCAAGCCT TGGCGCAGCA AATCCCTCAG CGTGAAGCAC
      TAACGTCATG GAGGAGCCGG TAGGGCGTCG GGCCACGTCG GTGGTTCGGA ACCGCGTCGT TTAGGGAGTC GCACTTCGTG
.............................................................................................
+2  S  A  T  V   S  M  L   S  V  K   P  P  G  P   E  A  P   R  P  T   P  E  A  M   K  P  A
 161  AGTGCCACGG TATCCATGCT CTCGGTCAAG CCTCCTGGGC CTGAGGCCCC CAGGCCCACA CCTGAAGCCA TGAAGCCGGC
      TCACGGTGCC ATAGGTACGA GAGCCAGTTC GGAGGACCCG GACTCCGGGG GTCCGGGTGT GGACTTCGGT ACTTCGGCCG
.............................................................................................
+2   P  N  N   Q  K  S  M   L  E  K   L  K  L   F  N  S  K   Q  G  S   K  A  G   E  G  P
 241  CCCCAACAAT CAGAAGTCCA TGCTGGAAAA GCTGAAACTT TTCAACAGTA AAGGGGGCTC AAAGGCAGGT GAGGGGCCGG
      GGGGTTGTTA GTCTTCAGGT ACGACCTTTT CGACTTTGAA AAGTTGTCAT TTCCCCCGAG TTTCCGTCCA CTCCCCGGCC
.............................................................................................
+2 G  S  R  D   T  S  C   E  R  L  E   T  L  P   S  F  E   E  S  E  E   L  E  A   A  S  R
 321  GGTCCCGGGA CACAAGCTGT GAGCGGCTGG AGACTCTGCC CAGCTTCGAA GAGAGCGAGG AGCTGGAGGC CGCCAGTCGC
      CCAGGGCCCT GTGTTCGACA CTCGCCGACC TCTGAGACGG GTCGAAGCTT CTCTCGCTCC TCGACCTCCG GCGGTCAGCG
.............................................................................................
+2  M  L  T  T   V  G  P   A  S  S   S  P  K  I   A  L  K   G  I  A   Q  R  T  F   S  R  A
 401  ATGCTCACCA CCGTGGGCCC TGCTTCCAGC AGCCCCAAGA TTGCACTCAA GGGCATTGCC CAGAGGACTT TTAGCCGGGC
      TACGAGTGGT GGCACCCGGG ACGAAGGTCG TCGGGGTTCT AACGTGAGTT CCCGTAACGG GTCTCCTGAA AATCGGCCCG
.............................................................................................
+2   L  T  N   K  K  S  S   L  K  G   N  E  K   E  K  E  K   Q  Q  R   E  K  D   K  E  K
 481  ACTGACCAAC AAGAAGAGTT CTCTGAAAGG CAATGAGAAA GAGAAGGAGA AACAACAGCG GGAGAAGGAT AAGGAGAAAA
      TGACTGGTTG TTCTTCTCAA GAGACTTTCC GTTACTCTTT CTCTTCCTCT TTGTTGTCGC CCTCTTCCTA TTCCTCTTTT
.............................................................................................
+2 S  K  D  L   A  K  R   A  S  V  T   E  R  L   D  L  K   E  E  P  K   E  D  P   S  G  A
 561  GCAAGGACCT TGCCAAGAGA GCCTCTGTGA CGGAGAGGCT GGACCTCAAG GAGGAGCCAA AAGAAGACCC CAGTGGAGCA
      CGTTCCTGGA ACGGTTCTCT CGGAGACACT GCCTCTCCGA CCTGGAGTTC CTCCTCGGTT TTCTTCTGGG GTCACCTCGT
.............................................................................................
+2  A  V  P  E   M  P  K   K  S  S   K  I  A  S   F  I  P   K  G  G   K  L  N  S   A  K  K
 641  GCTGTGCCCG AGATGCCAAA AAAGTCCTCC AAGATTGCCA GCTTCATCCC CAAAGGGGGG AAGCTCAACA GTGCCAAGAA
      CGACACGGGC TCTACGGTTT TTTCAGGAGG TTCTAACGGT CGAAGTAGGG GTTTCCCCCC TTCGAGTTGT CACGGTTCTT
.............................................................................................
+2   E  P  M   A  P  S  H   S  G  I   P  K  P   G  M  K  S   M  P  G   K  S  P   S  A  P
 721  GGAGCCCATG GCCCCTTCCC ACAGTGGAAT ACCAAAACCA GGAATGAAGA GCATGCCCGG GAAATCCCCA AGTGCCCCAG
      CCTCGGGTAC CGGGGAAGGG TGTCACCTTA TGGTTTTGGT CCTTACTTCT CGTACGGGCC CTTTAGGGGT TCACGGGGTC
.............................................................................................
+2 A  P  S  K   E  G  E   R  S  R  S   G  K  L   S  S  G   L  P  Q  Q   K  P  Q   L  D  G
 801  CGCCTTCCAA GGAAGGGGAG CGGAGCCGGA GTGGGAAGCT GAGCTCACGA CTCCCCCAGC AGAAGCCCCA GCTGGACGGC
      GCGGAAGGTT CCTTCCCCTC GCCTCGGCCT CACCCTTCGA CTCGAGTGCT GAGGGGGTCG TCTTCGGGGT CGACCTGCCG
.............................................................................................
+2  R  H  S  S   S  S  S   S  L  A   S  S  E  G   K  G  P   G  G  T   T  L  N  H   S  I  S
 881  AGACACTCCA GTTCCTCTTC CAGCCTGGCG TCCTCAGAAG GAAAAGGCCC AGGAGGGACC ACCCTGAACC ACAGCATCAG
      TCTGTGAGGT CAAGGAGAAG GTCGGACCGC AGGAGTCTTC CTTTTCCGGG TCCTCCCTGG TGGGACTTGG TGTCGTAGTC
.............................................................................................
+2   S  Q  T   V  S  G  S   V  G  T   T  Q  T   T  G  S  N   T  V  S   V  Q  L   P  Q  P
 961  CAGCCAGACT GTCAGTGGGT CTGTCGGGAC CACCCAGACC ACAGGAAGCA ATACCGTCAG TGTTCAGCTA CCTCAGCCCC
      GTCGGTCTGA CAGTCACCCA GACAGCCCTG GTGGGTCTGG TGTCCTTCGT TATGGCAGTC ACAAGTCGAT GGAGTCGGGG
.............................................................................................
+2 Q  Q  Q  Y   N  H  P   N  T  A  T   V  V  P   F  L  Y   R  S  Q  T   D  T  E   G  N  V
1041  AGCAGCAATA CAACCATCCC AACACTGCCA CGGTTGTACC TTTCCTGTAC AGGTCTCAGA CGGACACTGA AGGGAATGTT
      TCGTCGTTAT GTTGGTAGGG TTGTGACGGT GCCAACATGG AAAGGACATG TCCAGAGTCT GCCTGTGACT TCCCTTACAA
.............................................................................................
+2  T  A  E  S   S  S  T   G  V  S   V  E  P  S   H  F  T   K  T  G   Q  P  A  L   E  E  L
1121  ACTGCCGAGT CAAGCTCAAC AGGTGTGAGC GTGGAGCCCA GCCACTTCAC CAAGACTGGA CAGCCTGCTC TGGAAGAACT
      TGACGGCTCA GTTCGAGTTG TCCACACTCG CACCTCGGGT CGGTGAAGTG GTTCTGACCT GTCGGACGAG ACCTTCTTGA
.............................................................................................
```



```
+2   T  G  E   D  P  E  A   R  R  L   R  T  V   K  N  I  A   D  L  R   Q  N  L   E  E  T
1201  CACTGGGGAA GATCCTGAGG CTCGGCGGCT GCGGACAGTG AAGAACATCG CTGATCTGCG GCAGAATTTG GAGGAAACCA
      GTGACCCCTT CTAGGACTCC GAGCCGCCGA CGCCTGTCAC TTCTTGTAGC GACTAGACGC CGTCTTAAAC CTCCTTTGGT
.............................................................................................
+2 M  S  S  L   R  G  T   Q  V  T  K   P  N
                                     EcoRI
                                    ------
1281  TGTCCAGTTT AAGGGGAACT CAGGTTACAA AGCCGAATTC
      ACAGGTCAAA TTCCCCTTGA GTCCAATGTT TCGGCTTAAG
.............................................................................................
```

```
 +1                                                R A R A  A A R  A G A  K S G H  K R L
                 EcoRI
                 ------
    1  GATATCTGCA GAATTCGGCT TACTCACTAT AGGGCTCGAG CGGCCGCCCG AGCAGGTGCC|AAATCCGGCC ACAAGCGTCT
       CTATAGACGT CTTAAGCCGA ATGAGTGATA TCCCGAGCTC GCCGGCGGGC TCGTCCACGG TTTAGGCCGG TGTTCGCAGA
 ......................................................................................

 +1   I K D  L Q Q D  V T D  G V L  L A Q I  I Q V  V A N  E K I
   81  CATCAAGGAT CTCCAGCAAG ATGTGACAGA TGGCGTCCTC CTGGCCCAGA TTATCCAGGT TGTGGCAAAT GAAAAGATTG
       GTAGTTCCTA GAGGTCGTTC TACACTGTCT ACCGCAGGAG GACCGGGTCT AATAGGTCCA ACACCGTTTA CTTTTCTAAC
 ......................................................................................

 +1 E D I N  G C P  K N R S  Q M I  E N I  D A C L  N F L  A A K
  161  AAGACATCAA TGGCTGTCCG AAGAACAGAT CCCAAATGAT TGAAAACATA GATGCCTGCT TGAATTTCCT GGCAGCTAAG
       TTCTGTAGTT ACCGACAGGC TTCTTGTCTA GGGTTTACTA ACTTTTGTAT CTACGGACGA ACTTAAAGGA CCGTCGATTC
 ......................................................................................

 +1 G I N I  Q G L  S A E  E I R N  G N L  K A I  L G L F  F S L
  241  GGAATAAACA TCCAGGGGCT GTCTGCAGAA GAGATCAGGA ATGGAAACCT CAAGGCCATT CTAGGCCTCT TCTTCAGCCT
       CCTTATTTGT AGGTCCCCGA CAGACGTCTT CTCTAGTCCT TACCTTTGGA GTTCCGGTAA GATCCGGAGA AGAAGTCGGA
 ......................................................................................

 +1   S R Y  K Q Q Q  Q Q P  Q K Q  H L S S  P L P  P A V  S Q V
  321  CTCCCGATAC AAGCAGCAGC AGCAGCAGCC CCAGAAGCAG CACCTCTCCT CACCTCTGCC GCCCGCCGTA TCCCAGGTGG
       GAGGGCTATG TTCGTCGTCG TCGTCGTCGG GGTCTTCGTC GTGGAGAGGA GTGGAGACGG CGGGCGGCAT AGGGTCCACC
 ......................................................................................

 +1 A G A P  S Q C  Q A G T  P Q Q  Q V P  V T P Q  A P C  Q P H
  401  CCGGGGCCCC CTCCCAGTGC CAGGCTGGCA CCCCTCAGCA GCAGGTGCCA GTCACTCCCC AAGCCCCGTG CCAGCCTCAC
       GGCCCCGGGG GAGGGTCACG GTCCGACCGT GGGGAGTCGT CGTCCACGGT CAGTGAGGGG TTCGGGGCAC GGTCGGAGTG
 ......................................................................................

 +1 Q P A P  H Q Q  S K A  Q A E M  Q S R  L S G  P T A R  V S A
  481  CAGCCAGCGC CACATCAGCA GTCAAAAGCA CAAGCTGAAA TGCAGTCCAG ACTTTCAGGT CCTACCGCGA GGGTATCCGC
       GTCGGTCGCG GTGTAGTCGT CAGTTTTCGT GTTCGACTTT ACGTCAGGTC TGAAAGTCCA GGATGGCGCT CCCATAGGCG
 ......................................................................................

 +1   A G S  E A K T  R G G  S T T  A N N R  R S Q  S F N  N Y D
  561  TGCAGGCAGC GAGGCCAAAA CACGCGGAGG GTCAACTACT GCTAACAACC GACGCAGCCA GAGCTTTAAC AACTATGATA
       ACGTCCGTCG CTCCGGTTTT GTGCGCCTCC CAGTTGATGA CGATTGTTGG CTGCGTCGGT CTCGAAATTG TTGATACTAT
 ......................................................................................

 +1 K S K P  V T S  P P P P  P S S  H E K  E P L A  S S A  S S H
  641  AATCCAAACC AGTCACCTCC CCACCCCCAC CGCCAAGCAG CCACGAGAAA GAGCCTTTGG CAAGTTCAGC CTCCTCCCAC
       TTAGGTTTGG TCAGTGGAGG GGTGGGGGTG GCGGTTCGTC GGTGCTCTTT CTCGGAAACC GTTCAAGTCG GAGGAGGGTG
 ......................................................................................

 +1 P G M S  D N A  P A S  L E S G  S S S  T P T  N C S T  S S A
  721  CCCGGAATGA GTGACAATGC ACCTGCTTCC TTGGAGAGCG GCAGCAGCTC CACCCCTACT AATTGCAGTA CCTCCTCGGC
       GGGCCTTACT CACTGTTACG TGGACGAAGG AACCTCTCGC CGTCGTCGAG GTGGGGATGA TTAACGTCAT GGAGGAGCCG
 ......................................................................................

 +1   I P Q  P G A A  T K P  W R S  K S L S  V K H  S A T  V S M
  801  CATCCCGCAG CCCGGTGCAG CCACCAAGCC TTGGCGCAGC AAATCCCTCA GCGTGAAGCA CAGTGCCACG GTATCCATGC
       GTAGGGCGTC GGGCCACGTC GGTGGTTCGG AACCGCGTCG TTTAGGGAGT CGCACTTCGT GTCACGGTGC CATAGGTACG
 ......................................................................................

 +1 L S V K  P P G  P E A P  R P T  P E A  M K P A  P N N  Q K S
  881  TCTCGGTCAA GCCTCCTGGG CCTGAGGCCC CCAGGCCCAC ACCTGAAGCC ATGAAGCCGG CCCCCAACAA TCAGAAGTCC
       AGAGCCAGTT CGGAGGACCC GGACTCCGGG GGTCCGGGTG TGGACTTCGG TACTTCGGCC GGGGGTTGTT AGTCTTCAGG
 ......................................................................................

 +1 M L E K  L K L  F N S  K G G S  K A G  E G P  G S R D  T S C
  961  ATGCTGGAAA AGCTGAAACT TTTCAACAGT AAAGGGGGCT CAAAGGCAGG TGAGGGGCCG GGGTCCCGGG ACACAAGCTG
       TACGACCTTT TCGACTTTGA AAAGTTGTCA TTTCCCCCGA GTTTCCGTCC ACTCCCCGGC CCCAGGGCCC TGTGTTCGAC
 ......................................................................................

 +1   E R L  E T L P  S F E  E S E  E L E A  A S R  M L T  T V G
 1041  TGAGCGGCTG GAGACTCTGC CCAGCTTCGA AGAGAGCGAG GAGCTGGAGG CCGCCAGTCG CATGCTCACC ACCGTGGGCC
       ACTCGCCGAC CTCTGAGACG GGTCGAAGCT TCTCTCGCTC CTCGACCTCC GGCGGTCAGC GTACGAGTGG TGGCACCCGG
 ......................................................................................

 +1 P A S S  S P K  I A L K  G I A  Q R T  F S R A  L T N  K K S
 1121  CTGCTTCCAG CAGCCCCAAG ATTGCACTCA AGGGCATTGC CCAGAGGACT TTTAGCCGGG CACTGACCAA CAAGAAGAGT
       GACGAAGGTC GTCGGGGTTC TAACGTGAGT TCCCGTAACG GGTCTCCTGA AAATCGGCCC GTGACTGGTT GTTCTTCTCA
 ......................................................................................
```



```
 +1 S L K G  N E K  E K E  K Q Q R  E K D  K E K  S K D L  A K R
 1201  TCTCTGAAAG GCAATGAGAA AGAGAAGGAG AAACAACAGC GGGAGAAGGA TAAGGAGAAA AGCAAGGACC TTGCCAAGAG
       AGAGACTTTC CGTTACTCTT TCTCTTCCTC TTTGTTGTCG CCCTCTTCCT ATTCCTCTTT TCGTTCCTGG AACGGTTCTC
 ......................................................................................

 +1   A S V  T E R L  D L K  E E P  K E D P  S G A  A V P  E M P
 1281  AGCCTCTGTG ACGGAGAGGC TGGACCTCAA GGAGGAGCCA AAAGAAGACC CCAGTGGAGC AGCTGTGCCC GAGATGCCAA
       TCGGAGACAC TGCCTCTCCG ACCTGGAGTT CCTCCTCGGT TTTCTTCTGG GGTCACCTCG TCGACACGGG CTCTACGGTT
 ......................................................................................

 +1 K K S S  K I A  S F I P  K G G  K L N  S A K K  E P M  A P S
 1361  AAAAGTCCTC CAAGATTGCC AGCTTCATCC CCAAAGGGGG GAAGCTCAAC AGTGCCAAGA AGGAGCCCAT GGCCCCTTCC
       TTTTCAGGAG GTTCTAACGG TCGAAGTAGG GGTTTCCCCC CTTCGAGTTG TCACGGTTCT TCCTCGGGTA CCGGGGAAGG
 ......................................................................................

 +1   H S G I  P K P  G M K  S M P G  K S P  S A P  A P S K  E G E
 1441  CACAGTGGAA TACCAAAACC AGGAATGAAG AGCATGCCCG GGAAATCCCC AAGTGCCCCA GCGCCTTCCA AGGAAGGGGA
       GTGTCACCTT ATGGTTTTGG TCCTTACTTC TCGTACGGGC CCTTTAGGGG TTCACGGGGT CGCGGAAGGT TCCTTCCCCT
 ......................................................................................

 +1   R S R  S G K L  S S G  L P Q  Q K P Q  L D G  R H S  S S S
 1521  GCGGAGCCGG AGTGGGAAGC TGAGCTCAGG ACTCCCCCAG CAGAAGCCCC AGCTGGACGG CAGACACTCC AGTTCCTCTT
       CGCCTCGGCC TCACCCTTCG ACTCGAGTCC TGAGGGGGTC GTCTTCGGGG TCGACCTGCC GTCTGTGAGG TCAAGGAGAA
 ......................................................................................

 +1 S S L A  S S E  G K G P  G G T  T L N  H S I S  S Q T  V S G
 1601  CCAGCCTGGC GTCCTCAGAA GGAAAAGGCC CAGGAGGGAC CACCCTGAAC CACAGCATCA GCAGCCAGAC TGTCAGTGGG
       GGTCGGACCG CAGGAGTCTT CCTTTTCCGG GTCCTCCCTG GTGGGACTTG GTGTCGTAGT CGTCGGTCTG ACAGTCACCC
 ......................................................................................

 +1   S V G T  T Q T  T G S  N T V S  V Q L  P Q P  Q Q Q Y  N H P
 1681  TCTGTCGGGA CCACCCAGAC CACAGGAAGC AATACCGTCA GTGTTCAGCT ACCTCAGCCC CAGCAGCAAT ACAACCATCC
       AGACAGCCCT GGTGGGTCTG GTGTCCTTCG TTATGGCAGT CACAAGTCGA TGGAGTCGGG GTCGTCGTTA TGTTGGTAGG
 ......................................................................................

 +1   N T A  T V A P  F L Y  R S Q  T D T E  G N V  T A E  S S S
 1761  CAACACTGCC ACGGTTGCAC CTTTCCTGTA CAGGTCTCAG ACGGACACTG AAGGGAATGT TACTGCCGAG TCAAGCTCAA
       GTTGTGACGG TGCCAACGTG GAAAGGACAT GTCCAGAGTC TGCCTGTGAC TTCCCTTACA ATGACGGCTC AGTTCGAGTT
 ......................................................................................

 +1 T G V S  V E P  S H F T  K T G  Q P A  L E E L  T G E  D P E
 1841  CAGGTGTGAG CGTGGAGCCC AGCCACTTCA CCAAGACTGG ACAGCCTGCT CTGGAAGAAC TCACTGGGGA AGATCCTGAG
       GTCCACACTC GCACCTCGGG TCGGTGAAGT GGTTCTGACC TGTCGGACGA GACCTTCTTG AGTGACCCCT TCTAGGACTC
 ......................................................................................

 +1 A R R L  R T V  K N I  A D L R  Q N L  E E T  M S S L  R G A
 1921  GCTCGGCGGC TGCGGACAGT GAAGAACATC GCTGATCTGC GGCAGAATTT GGAGGAAACC ATGTCCAGTT TAAGGGGAGC
       CGAGCCGCCG ACGCCTGTCA CTTCTTGTAG CGACTAGACG CCGTCTTAAA CCTCCTTTGG TACAGGTCAA ATTCCCCTCG
 ......................................................................................

 +1   Q V T  K P N
                   EcoRI
                  -------
 2001  TCAGGTTACA|AAGCCGAATT C
       AGTCCAATGT TTCGGCTTAA G
 ......................................................................................
```

```
11/24/97 04:11PM          D2.1-5  (hh2UNC53  5'RACE  02/04/97)                    Page 1

 +3                             S L  *  G  S  S   G  R  P   G  R  C  S   S  T  P   T  N  C
               EcoRI
               ------
   1  GATATCTGCA GAATTCGGCT TACTCACTAT AGGGCTCGAG CGGCCGCCCG GGCAGGTGCA GCTCCACCCC TACTAATTGC
      CTATAGACGT CTTAAGCCGA ATGAGTGATA TCCCGAGCTC GCCGGCGGGC CCGTCCACGT CGAGGTGGGG ATGATTAACG
.........................................................................................
 +3  S  T  S  S   A  I  P   Q  P  G   A  A  T  K   P  W  R   S  K  S   L  S  V  K   H  S  A
  81  AGTACCTCCT CGGCCATCCC GCAGCCCGGT GCAGCCACCA AGCCTTGCCG CAGCAAATCC CTCAGCGTGA AGCACAGTGC
      TCATGGAGGA GCCGGTAGGG CGTCGGGCCA CGTCGGTGGT TCGGAACGGC GTCGTTTAGG GAGTCGCACT TCGTGTCACG
.........................................................................................
 +3   T  V  S   M  L  S  V   K  P  P   G  P  E   A  P  P  P   T  P  E   A  M  K   P  A  P
 161  CACGGTATCC ATGCTCTCGG TCAAGCCTCC TGGGCCTGAG GCCCCCACGC CCACACCTGA AGCCATGAAG CCGGCCCCCA
      GTGCCATAGG TACGAGAGCC AGTTCGGAGG ACCCGGACTC CGGGGGTGCG GGTGTGGACT TCGGTACTTC GGCCGGGGGT
.........................................................................................
 +3 N  N  Q  K   S  M  L   E  K  L  K   L  F  N   S  K  G   G  S  K  A   G  E  G   P  G  S
 241  ACAATCAGAA GTCCATGCTG GAAAAGCTGA AACTTTTCAA CAGTAAAGGG GGCTCAAAGG CAGGTGAGGG GCCGGGGTCC
      TGTTAGTCTT CAGGTACGAC CTTTTCGACT TTGAAAAGTT GTCATTTCCC CCGAGTTTCC GTCCACTCCC CGGCCCCAGG
.........................................................................................
 +3  R  D  T  S   C  E  R   L  E  T   L  P  S  F   E  E  S   E  E  L   E  A  A  S   R  M  L
 321  CGGGACACAA GCTGTGAGCG GCTGGAGACT CTGCCCAGCT TCGAAGAGAG CGAGGAGCTG GAGGCCGCCA GTCGCATGCT
      GCCCTGTGTT CGACACTCGC CGACCTCTGA GACGGGTCGA AGCTTCTCTC GCTCCTCGAC CTCCGGCGGT CAGCGTACGA
.........................................................................................
 +3   T  T  V   G  P  A  S   S  S  P   K  I  A   L  K  G  I   A  Q  R   T  F  S   R  A  L
 401  CACCACCGTG GGCCCTGCTT CCAGCAGCCC CAAGATTGCA CTCAAGGGCA TTGCCCAGAG GACTTTTAGC CGGGCACTGA
      GTGGTGGCAC CCGGGACGAA GGTCGTCGGG GTTCTAACGT GAGTTCCCGT AACGGGTCTC CTGAAAATCG GCCCGTGACT
.........................................................................................
 +3 T  N  K  K   S  S  L   K  G  N  E   K  E  K   E  K  Q   Q  R  E  K   D  K  E   K  S  K
 481  CCAACAAGAA GAGTTCTCTG AAAGGCAATG AGAAAGAGAA GGAGAAACAA CAGCGGGAGA AGGATAAGGA GAAAAGCAAG
      GGTTGTTCTT CTCAAGAGAC TTTCCGTTAC TCTTTCTCTT CCTCTTTGTT GTCGCCCTCT TCCTATTCCT CTTTTCGTTC
.........................................................................................
 +3  D  L  A  K   R  A  S   V  T  E   R  L  D  L   K  E  E   P  K  E   D  P  S  G   A  A  V
 561  GACCTTGCCA AGAGAGCCTC TGTGACGGAG AGGCTGGACC TCAAGGAGGA GCCAAAAGAA GACCCCAGTG GAGCAGCTGT
      CTGGAACGGT TCTCTCGGAG ACACTGCCTC TCCGACCTGG AGTTCCTCCT CGGTTTTCTT CTGGGGTCAC CTCGTCGACA
.........................................................................................
 +3   P  E  M   P  K  K  S   S  K  I   A  S  F   I  P  K  G   G  K  L   N  S  A   K  K  E
 641  GCCCGAGATG CCAAAAAAGT CCTCCAAGAT TGCCAGCTTC ATCCCCAAAG GGGGGAAGCT CAACAGTGCC AAGAAGGAGC
      CGGGCTCTAC GGTTTTTTCA GGAGGTTCTA ACGGTCGAAG TAGGGGTTTC CCCCCTTCGA GTTGTCACGG TTCTTCCTCG
.........................................................................................
 +3 P  M  A  P   S  H  S   G  I  P  K   P  G  M   K  S  M   P  G  K  S   P  S  A   P  A  P
 721  CCATGGCCCC TTCCCACAGT GGAATACCAA AACCAGGAAT GAAGAGCATG CCCGGGAAAT CCCCAAGTGC CCCAGCGCCT
      GGTACCGGGG AAGGGTGTCA CCTTATGGTT TTGGTCCTTA CTTCTCGTAC GGGCCCTTTA GGGGTTCACG GGGTCGCGGA
.........................................................................................
 +3  S  K  E  G   E  R  S   R  S  G   K  L  S  S   G  L  P   Q  Q  K   P  Q  L  D   G  R  H
 801  TCCAAGGAAG GGGAGCGGAG CCGGAGTGGG AAGCTGAGCT CAGGACTCCC CCAGCAGAAG CCCCAGCTGG ACGGCAGACA
      AGGTTCCTTC CCCTCGCCTC GGCCTCACCC TTCGACTCGA GTCCTGAGGG GGTCGTCTTC GGGGTCGACC TGCCGTCTGT
.........................................................................................
 +3   S  S  S   S  S  S  L   A  S  S   E  G  K   G  P  G  G   T  T  L   N  H  S   I  S  S
 881  CTCCAGTTCC TCTTCCAGCC TGGCGTCCTC AGAAGGAAAA GGCCCAGGAG GGACCACCCT GAACCACAGC ATCAGCAGCC
      GAGGTCAAGG AGAAGGTCGG ACCGCAGGAG TCTTCCTTTT CCGGGTCCTC CCTGGTGGGA CTTGGTGTCG TAGTCGTCGG
.........................................................................................
 +3 Q  T  V  S   G  S  V   G  T  T  Q   T  T  G   S  N  T   V  S  V  Q   L  P  Q   P  Q  Q
 961  AGACTGTCAG TGGGTCTGTC GGGACCACCC AGACCACAGG AAGCAATACC GTCAGTGTTC AGCTACCTCA GCCCCAGCAG
      TCTGACAGTC ACCCAGACAG CCCTGGTGGG TCTGGTGTCC TTCGTTATGG CAGTCACAAG TCGATGGAGT CGGGGTCGTC
.........................................................................................
 +3  Q  Y  N  H   P  N  T   A  T  V   A  P  F  L   Y  R  S   Q  T  D   T  E  G  N   V  T  A
1041  CAATACAACC ATCCCAACAC TGCCACGGTT GCACCTTTCC TGTACAGGTC TCAGACGGAC ACTGAAGGGA ATGTTACTGC
      GTTATGTTGG TAGGGTTGTG ACGGTGCCAA CGTGGAAAGG ACATGTCCAG AGTCTGCCTG TGACTTCCCT TACAATGACG
.........................................................................................
 +3   E  S  S   S  T  G  V   S  V  E   P  S  H   F  T  K  T   G  Q  P   A  L  E   E  L  T
1121  CGAGTCAAGC TCAACAGGTG TGAGCGTGGA GCCCAGCCAC TTCACCAAGA CTGGACAGCC TGCTCTGGAA GAACTCACTG
      GCTCAGTTCG AGTTGTCCAC ACTCGCACCT CGGGTCGGTG AAGTGGTTCT GACCTGTCGG ACGAGACCTT CTTGAGTGAC
.........................................................................................
```

```
11/24/97 04:11PM          D2.1-5  (hh2UNC53  5'RACE  02/04/97)                    Page 2

 +3 G  E  D  P   E  A  R   R  L  R  T   V  K  N   I  A  D   L  R  Q  N   L  E  E   T  V  S
1201  GGGAAGATCC TGAGGCTCGG CGGCTGCGGA CAGTGAAGAA CATCGCTGAT CTGCGGCAGA ATTTGGAGGA AACCGTGTCC
      CCCTTCTAGG ACTCCGAGCC GCCGACGCCT GTCACTTCTT GTAGCGACTA GACGCCGTCT TAAACCTCCT TTGGCACAGG
.........................................................................................
 +3  S  L  R  G   T  Q  V   T  K  P   N
                                    EcoRI
                                   -------
1281  AGTTTAAGGG GAACTCAGGT TACAAAGCCG AATTC
      TCAAATTCCC CTTGAGTCCA ATGTTTCGGC TTAAG
.........................................................................................
```

```
11/24/97 04:05PM        UNC53hh2_5'race_D4.1.1_CON.seq        Page 1

 +3                                     S L  *  C S S  G R P  G R S V  R D R  L E E
              EcoRI
              ------
   1  GATATCTGCA GAATTCGGCT TACTCACTAT AGGGCTCGAG CGGCCGCCCG GGCAGGTCAG TGCGAGACCG GCTGGAGGAG
      CTATAGACGT CTTAAGCCGA ATGAGTGATA TCCCGAGCTC GCCGGCGGGC CCGTCCAGTC ACGCTCTGGC CGACCTCCTC
.......................................................................................
 +3  A A P R  S R C  G R R  Y G S Y  T M K  E D D  S P G G  E P
  81  GCGGCTCCGC GAAGCCGGTG TGGGCGCAGG TATGGGTCTT ACACCATGAA GGAAGATGAC AGCCCAGGAG GGGAGCCACC
      CGCCGAGGCG CTTCGGCCAC ACCCGCGTCC ATACCCAGAA TGTGGTACTT CCTTCTACTG TCGGGTCCTC CCCTCGGTGG
.......................................................................................
 +3   L * *  T L A G  A M R  G G F  P R S V  C S D  L W E  S G A
 161  GCTGTGATGA ACGCTTGCAG GAGCCATGAG GGGCGGCTTT CCGCGCAGTG TTTGCAGTGA TTTATGGGAG TCAGGGGCTG
      CGACACTACT TGCGAACGTC CTCGGTACTC CCCGCCGAAA GGCGCGTCAC AAACGTCACT AAATACCCTC AGTCCCCGAC
.......................................................................................
 +3 A R E G  C G C  * Q L G  G L N  A H L  H C Y *  H I N  Y L C
 241  CCAGAGAGGG ATGTGGATGT TAGCAGCTTG GTGGATTAAA TGCCCATTTA CATTGTTATT GACATATAAA TTATCTGTGC
      GGTCTCTCCC TACACCTACA ATCGTCGAAC CACCTAATTT ACGGGTAAAT GTAACAATAA CTGTATATTT AATAGACACG
.......................................................................................
 +3  Q G D S  E G K  G F W  G S N V  V C A  S G Y  G L V F  A Q
 321  CAAGGAGACA GTGAAGGAAA GGGTTTCTGG GGCAGTAATG TGGTATGTGC CTCAGGATAT GGACTTGTCT TTGCCCAGAT
      GTTCCTCTGT CACTTCCTTT CCCAAAGACC CCGTCATTAC ACCATACACG GAGTCCTATA CCTGAACAGA AACGGGTCTA
.......................................................................................
 +3   C P L  G S H E  A S L  K P K  V K P Q  D L A  L M M  P A A
 401  TTGTCCTCTG GGCTCTCATG AAGCGAGTCT CAAACCTAAA GTAAAACCCC AAGACTTAGC TCTCATGATG CCGGCAGCAG
      AACAGGAGAC CCGAGAGTAC TTCGCTCAGA GTTTGGATTT CATTTTGGGG TTCTGAATCG AGAGTACTAC GGCCGTCGTC
.......................................................................................
 +3 V T L I  A K A  V V P S  E V P  E W I  L R R Q  E D L  * R G
 481  TCACCCTCAT TGCAAAGGCA GTGGTGCCGA GTGAGGTTCC CGAGTGGATT TTAAGAAGAC AGGAAGACCT TTGAAGAGGT
      AGTGGGAGTA ACGTTTCCGT CACCACGGCT CACTCCAAGG GCTCACCTAA AATTCTTCTG TCCTTCTGGA AACTTCTCCA
.......................................................................................
 +3  G A D F  L G W  R E L  C L A V  A C I  T F V  W V I L  F L
 561  GGTGCTGATT TCCTTGGCTG GCGGGAACTC TGTCTGGCTG TTGCATGCAT CACTTTTGTG TGGGTTATTT TGTTCCTCTG
      CCACGACTAA AGGAACCGAC CGCCCTTGAG ACAGACCGAC AACGTACGTA GTGAAAACAC ACCCAATAAA ACAAGGAGAC
.......................................................................................
 +3   G F G  S I A E  G E R  G F Y  F W E M  E S V  S E S  S Q Q
 641  TGGATTTGGA AGCATCGCTG AAGGAGAGAG AGGATTTTAT TTCTGGGAAA TGGAATCGGT TTCTGAGTCC AGCCAACAGC
      ACCTAAACCT TCGTAGCGAC TTCCTCTCTC TCCTAAAATA AAGACCCTTT ACCTTAGCCA AAGACTCAGG TCGGTTGTCG
.......................................................................................
 +3 Q K R K  P V I  H G L E  D Q K  R I Y  T D W A  N H Y  L A K
 721  AGAAGAGAAA GCCAGTTATC CACGGACTGG AAGATCAAAA GAGGATCTAC ACAGACTGGG CCAATCATTA CCTAGCCAAA
      TCTTCTCTTT CGGTCAATAG GTGCCTGACC TTCTAGTTTT CTCCTAGATG TGTCTGACCC GGTTAGTAAT GGATCGGTTT
.......................................................................................
 +3  S G H K  R L I  R D L  Q Q D V  T D G  V L L  A Q I I  Q V
 801  TCCGGCCACA AGCGTCTCAT CAGGGATCTC CAGCAAGATG TGACAGATGG CGTCCTCCTG GCCCAGATTA TCCAGGTTGT
      AGGCCGGTGT TCGCAGAGTA GTCCCTAGAG GTCGTTCTAC ACTGTCTACC GCAGGAGGAC CGGGTCTAAT AGGTCCAACA
.......................................................................................
 +3   A N E  K I E D  I N G  C P K  N R S Q  M I E  N I D  A C L
 881  GGCAAATGAA AAGATTGAAG ACATCAATGG CTGTCCGAAG AACAGATCCC AAATGATTGA AAACATAGAT GCCTGCTTGA
      CCGTTTACTT TTCTAACTTC TGTAGTTACC GACAGGCTTC TTGTCTAGGG TTTACTAACT TTTGTATCTA CGGACGAACT
.......................................................................................
 +3 N F L A  A K G  I N I Q  G L S  A E E  I R N G  N L K  A I L
 961  ATTTCCTGGC AGCTAAGGGA ATAAACATCC AGGGGCTGTC TGCAGAAGAG ATCAGGAATG GAAACCTCAA GGCCATTCTA
      TAAAGGACCG TCGATTCCCT TATTTGTAGG TCCCCGACAG ACGTCTTCTC TAGTCCTTAC CTTTGGAGTT CCGGTAAGAT
.......................................................................................
 +3  G L F F  S L S  R Y K  Q Q Q Q  Q P Q  K Q H  L S S P  L P
1041  GGCCTCTTCT TCAGCCTCTC CCGATACAAG CAGCAGCAGC AGCAGCCCCA GAAGCAGCAC CTCTCCTCAC CTCTGCCGCC
      CCGGAGAAGA AGTCGGAGAG GGCTATGTTC GTCGTCGTCG TCGTCGGGGT CTTCGTCGTG GAGAGGAGTG GAGACGGCGG
.......................................................................................
 +3   A V S  Q V A G  A P S  Q C Q  A G T P  Q Q Q  V P V  T P Q
1121  CGCCGTATCC CAGGTGGCCG GGGCCCCCTC CCAGTGCCAG GCTGGCACCC CTCAGCAGCA GGTGCCAGTC ACTCCCCAAG
      GCGGCATAGG GTCCACCGGC CCCGGGGGAG GGTCACGGTC CGACCGTGGG GAGTCGTCGT CCACGGTCAG TGAGGGGTTC
.......................................................................................
```

```
11/24/97 04:05PM        UNC53hh2_5'race_D4.1.1_CON.seq        Page 2

 +3 A P C Q  P H Q  P A P H  Q Q S  K A Q  A E M Q  S R L  S G K
1201  CCCCGTGCCA GCCTCACCAG CCAGCACCAC ATCAGCAGTC AAAAGCACAA GCTGAAATGC AGTCCAGACT TTCAGGTAAG
      GGGGCACGGT CGGAGTGGTC GGTCGTGGTG TAGTCGTCAG TTTTCGTGTT CGACTTTACG TCAGGTCTGA AAGTCCATTC
.......................................................................................
 +3  P N
       EcoRI
      ------
1281  CCGAATTC
      GGCTTAAG
.......................................................................................
```

Fig. 11f.

11/24/97 04:17PM UNC53hh2_5'race_J4.1.4_CON.seq P Page 1

```
+2                          *  G  S  S   G  R  P   G  R  Y  G   A  C  A   D  P  Q   A  L  R
   1  GAATTCGGCT TACTCACTAT AGGGCTCGAG CGGCCGCCCG GGCAGGTACG GAGCATGTGC AGACCCGCAG GCGCTAAGGT
      CTTAAGCCGA ATGAGTGATA TCCCGAGCTC GCCGGCGGGC CCGTCCATGC CTCGTACACG TCTGGGCGTC CGCGATTCCA
.........................................................................................
+2 C  S  Q  S   G  G  L   L  G  S  G   D  C  H   P  K  R   D  L  S  G   A  L  A   A  P  L
  81  GCTCTCAGTC CGGGGGCCTT CTGGGCAGCG GTGACTGTCA CCCCAAGAGA GACCTTTCGG GGGCTCTTGC CGCACCTCTG
      CGAGAGTCAG GCCCCCGGAA GACCCGTCGC CACTGACAGT GGGGTTCTCT CTGGAAAGCC CCCGAGAACG GCGTGGAGAC
.........................................................................................
+2  L  R  R  G   *  H  C   G  A  R   E  Q  G  G   S  C  L   S  V  E   T  S  W  D   P  C
 161  CTGCGTCGCG GGTGACACTG CGGTGCTCGC GAGCAGGGTG GCAGCTGCCT CTCGGTGGAG ACGTCTTGGG ACCCTTGCGC
      GACGCAGCGC CCACTGTGAC GCCACGAGCG CTCGTCCCAC CGTCGACGGA GAGCCACCTC TGCAGAACCC TGGGAACGCG
.........................................................................................
+2   L  L  L   S  F  L  R   C  C  L   L  S  F   L  G  C  S   L  F  L   S  P  A   Q  T  R
 241  CCTTCTTCTC TCCTTCCTTC GCTGCTGTCT CCTTTCCTTC CTTGGCTGCT CGCTCTTTCT CTCGCCGGCT CAGACCCGTA
      GGAAGAAGAG AGGAAGGAAG CGACGACAGA GGAAAGGAAG GAACCGACGA GCGAGAAAGA GAGCGGCCGA GTCTGGGCAT
.........................................................................................
+2 S  L  R  N   G  T  R   G  S  P  P   P  L  S   L  P  K   G  P  P  H   F  G  D   A  V  T
 321  GCCTCCGGAA CGGGACTCGT GGGTCGCCGC CGCCTCTGTC TCTTCCAAAG GGGCCTCCTC ACTTCGGAGA TGCAGTGACA
      CGGAGGCCTT GCCCTGAGCA CCCAGCGGCG GCGGAGACAG AGAAGGTTTC CCCGGAGGAG TGAAGCCTCT ACGTCACTGT
.........................................................................................
+2  S  *  Y  G   R  P  S   L  C  F   P  G  G  N   L  Q  E   A  A  A   P  E  R  P   E  L
 401  AGTTAATATG GGCGTCCAAG CCTCTGTTTC CCAGGAGGAA ATTTGCAAGA GGCGGCAGCC CCTGAGCGCC CAGAGCTCTT
      TCAATTATAC CCGCAGGTTC GGAGACAAAG GGTCCTCCTT TAAACGTTCT CCGCCGTCGG GGACTCGCGG GTCTCGAGAA
.........................................................................................
+2   K  G  H   P  G  E  V   *  D  P   A  A  A   S  V  Q  V   G  P  A   E  R  R   G  Q  S
 481  GAAAGGCCAC CCAGGAGAGG TGTGAGACCC GGCGGCAGCA TCCGTCCAGG TGGGACCCGC TGAGCGCCGT GGCCAGTCCC
      CTTTCCGGTG GGTCCTCTCC ACACTCTGGG CCGCCGTCGT AGGCAGGTCC ACCCTGGGCG ACTCGCGGCA CCGGTCAGGG
.........................................................................................
+2 P  I  P  I   P  A  P   Q  R  R  A   R  P  I   A  S  S   S  P  T  W   G  F  F   F  L  A
 561  CCATTCCCAT CCCGGCACCC CAAAGGCGCG CTCGCCCGAT TGCTTCGAGT TCCCCGACCT GGGGATTTTT TTTTTTAGCC
      GGTAAGGGTA GGGCCGTGGG GTTTCCGCGC GAGCGGGCTA ACGAAGCTCA AGGGGCTGGA CCCCTAAAAA AAAAAATCGG
.........................................................................................
+2  A  G  G  G   R  L  V   G  *  G   P  A  P  A   L  L  P   A  L  P   L  A  V  A   P  A
 641  GCTGGTGGTG GGCGCCTCGT GGGCTAAGGC CCGGCGCCTG CTCTGCTACC CGCGCTGCCT TTAGCGGTCG CCCCCGCCGC
      CGACCACCAC CCGCGGAGCA CCCGATTCCG GGCCGCGGAC GAGACGATGG GCGCGACGGA AATCGCCAGC GGGGGCGGCG
.........................................................................................
+2   A  A  R   D  V  L  G   K  P  K   P  R  E   K  M  P  A   I  L  V   A  S  K   M  K  S
 721  CGCTGCCAGG GACGTGCTGG GAAAGCCCAA GCCCCGGGAG AAGATGCCGG CCATCCTGGT CGCCTCCAAA ATGAAGTCGG
      GCGACGGTCC CTGCACGACC CTTTCGGGTT CGGGGCCCTC TTCTACGGCC GGTAGGACCA GCGGAGGTTT TACTTCAGCC
.........................................................................................
+2 G  L  P  K   P  V  H   S  A  A  P   I  L  H   V  P  P   A  R  A  G   P  Q  P   C  Y  L
 801  GACTGCCCAA ACCCGTGCAC AGCGCCGCGC CCATCCTGCA CGTGCCCCCG GCCCGGGCGG GCCCCCAGCC CTGCTACCTG
      CTGACGGGTT TGGGCACGTG TCGCGGCGCG GGTAGGACGT GCACGGGGGC CGGGCCCGCC CGGGGGTCGG GACGATGGAC
.........................................................................................
+2  K  L  G  S   K  V  E   V  S  K   T  T  Y  P   S  Q  I   P  L  K   S  Q  V  L   Q  G
 881  AAGTTGGGAA GCAAGGTGGA GGTGAGCAAG ACCACCTATC CTAGCCAGAT CCCCCTGAAA TCGCAGGTGC TGCAGGGGCT
      TTCAACCCTT CGTTCCACCT CCACTCGTTC TGGTGGATAG GATCGGTCTA GGGGGACTTT AGCGTCCACG ACGTCCCCGA
.........................................................................................
+2   Q  E  P   A  G  E  G   L  P  L   R  K  S   G  S  V  E   N  G  F   D  T  Q   I  Y  T
 961  GCAGGAGCCA GCGGGGGAGG GGCTCCCGCT GCGGAAGAGC GGCTCGGTGG AAAACGGGTT CGATACCCAG ATCTACACAG
      CGTCCTCGGT CGCCCCCTCC CCGAGGGCGA CGCCTTCTCG CCGAGCCACC TTTTGCCCAA GCTATGGGTC TAGATGTGTC
.........................................................................................
+2 D  W  A  N   H  Y  L   A  K  S  G   H  K  R   L  I  K   D  L  Q  Q   E  S  R   I  L  Q
1041  ACTGGGCCAA TCATTACCTA GCCAAATCCG GCCACAAGCG TCTCATCAAG GATCTCCAGC AAGAAAGCCG AATTCTGCAG
      TGACCCGGTT AGTAATGGAT CGGTTTAGGC CGGTGTTCGC AGAGTAGTTC CTAGAGGTCG TTCTTTCGGC TTAAGACGTC
.........................................................................................
+2  I  U
1121  ATAC
      TATG
.........................................................................................
```

```
11/24/97 04:25PM              hh2UNC53_G4.1.1+M13R.          G         Page 1

  1  CTTACAAAGN TGTTGAGGCA ACGGNTTTNT GCTCCAGGCC CGGAGACAGG CTTGTTTGGG AGNTTGNCCA GCTTCCCATT
     GAATGTTTCN ACAACTCCGT TGCCNAAANA CGAGGTCCGG GCCTCTGTCC GAACAAACCC TCNAACNGGT CGAAGGGTAA
..........................................................................................
 +3                                                                 |P  F  P  R  D  T  G
 81  GCCCCGGACA ACCAGGGCAG TGGCCTCAGA GCAGAACCAG NTCGNTGGGT TTGTCTTGCC|CCCTTCCCAA GGGACACAGG
     CGGGGCCTGT TGGTCCCGTC ACCGGAGTCT CGTCTTGGTC NAGCNACCCA AACAGAACGG GGGAAGGGTT CCCTGTGTCC
..........................................................................................
 +3  R  Q  T  L  F  A  E  T  S  T  A  F  I  I  T  K  P  L  P  C  S  D  N  E  F  Q
161  CAGACAGACA CTGTTTGCTG AGACTTCCAC AGCTTTCATT ATTACAAAAC CTCTGCCATG CTCAGACAAT GAGTTTCAGC
     GTCTGTCTGT GACAAACGAC TCTGAAGGTG TCGAAAGTAA TAATGTTTTG GAGACGGTAC GAGTCTGTTA CTCAAAGTCG
..........................................................................................
 +3 Q  M  M  A  L  L  P  R  A  S  V  E  H  L  E  *  Q  L  D  F  I  E  C  L  P  C  T
241  AAATGATGGC ACTACTTCCG AGAGCATCTG TAGAGCACCT AGAATAACAA CTGGACTTTA TTGAGTGTTT ACCATGCACC
     TTTACTACCG TGATGAAGGC TCTCGTAGAC ATCTCGTGGA TCTTATTGTT GACCTGAAAT AACTCACAAA TGGTACGTGG
..........................................................................................
 +3  K  P  W  A  K  H  F  I  C  R  L  F  V  F  T  A  N  P  V  G  R  Y  N  Y  P  H  S
321  AAGCCCTGGG CTAAACACTT CATCTGCAGG CTGTTCGTCT TTACGGCAAA CCCAGTAGGT AGGTATAACT ATCCCCACTC
     TTCGGGACCC GATTTGTGAA GTAGACGTCC GACAAGCAGA AATGCCGTTT GGGTCATCCA TCCATATTGA TAGGGGTGAG
..........................................................................................
 +3  A  D  A  E  T  E  A  Q  S  V  L  V  A  K  Q  A  H  Q  E  A  R  R  W  P  H  L
401  TGCAGATGCA GAAACGGAGG CACAGAGTGT TTTGGTAGCT AAACAAGCTC ACCAGGAGGC TAGAAGGTGG CCACACCTAG
     ACGTCTACGT CTTTGCCTCC GTGTCTCACA AAACCATCGA TTTGTTCGAG TGGTCCTCCG ATCTTCCACC GGTGTGGATC
..........................................................................................
 +3 A  G  P  P  D  S  T  N  C  L  P  L  L  C  C  H  Q  E  C  D  S  K  F  F  L  P  S
481  CTGGCCCCCC TGACTCCACC AACTGCCTCC CTTTGCTGTG TTGCATGCAA GAATGTGACT CCAAGTTTTT CCTTCCTTCT
     GACCGGGGGG ACTGAGGTGG TTGACGGAGG GAAACGACAC AACGTACGTT CTTACACTGA GGTTCAAAAA GGAAGGAAGA
..........................................................................................
 +3  G  S  N  S  G  F  T  L  L  S  N  Q  I  Y  T  D  W  A  N  H  Y  L  A  K  S  G  H
561  GGATCCAACT CTGGCTTCAC TCTGCTCAGC AACCAGATCT ACACAGACTG GGCCAATCAT TACCTAGCCA AATCCGGCCA
     CCTAGGTTGA GACCGAAGTG AGACGAGTCG TTGGTCTAGA TGTGTCTGAC CCGGTTAGTA ATGGATCGGT TTAGGCCGGT
..........................................................................................
 +3  K  R  L  I  K  D  L  Q  Q| E  S  R  I  P  A  H  W  R  P  L  L  V  D  P  S  S
641  CAAGCGTCTC ATCAAGGATC TCCAGCAA|GA AAGCCGAATT CCAGCACACT GGCGGCCGTT ACTAGTGGAT CCGAGCTCGG
     GTTCGCAGAG TAGTTCCTAG AGGTCGTTCT TTCGGCTTAA GGTCGTGTGA CCGCCGGCAA TGATCACCTA GGCTCGAGCC
..........................................................................................
 +3 V  P  K  L  G  V  I  M  V  I  V
721  TACCCAAGCT TGGCGTAATC ATGGTCATAG TGT
     ATGGGTTCGA ACCGCATTAG TACCAGTATC ACA
..........................................................................................
```

```
11/25/97 03:44PM          hh2UNC_F4_1_2+M13.Seq          R          Page 1

  1  NNGAAGAAGA GGGGGTAAAA AANNGTNTTN AANAAAANGG GATCCCTTTG NGGGGGTNTT TTTTTNGCCC CCCNGGNTTG
     NNCTTCTTCT CCCCCATTTT TTNNCANAAN TTNTTTTNCC CTAGGGAAAC NCCCCCANAA AAAAANCGGG GGGNCCNAAC
.........................................................................................
 81  TAAAAGGGGG TCCCTTANGG ATTNTTNANG GAAAGGCAAG GAAGGGAACC ACCTTTTCTT TGGGGNGGGN GTGAANACCG
     ATTTTCCCCC AGGGAATNCC TAANAANTNC CTTTCCGTTC CTTCCCTTGG TGGAAAAGAA ACCCCNCCCN CACTTNTGGC
.........................................................................................
161  NCACCACCCT GGGGGNGGAC NGGCAACCAN TNTTGCTTTT AANAAAGGGA ANTTNGGAAG AAAAGGGNGG TTTGGNTAAG
     NGTGGTGGGA CCCCCNCCTG NCCGTTGGTN ANAACGAAAA TTNTTTCCCT TNAANCCTTC TTTTCCCNCC AAACCNATTC
.........................................................................................
241  NTGGGGANGT TTTCGGGNCC AATNACAATG ACAGGTANGG TGGGATCCGC CGTAACNGGG GGGNAATGTC NNGGGTCCCT
     NACCCCTNCA AAAGCCCNGG TTANTGTTAC TGTCCATNCC ACCCTAGGCG GCATTGNCCC CCCNTTACAG NNCCCAGGGA
.........................................................................................
321  AAGTNTGGGG TTNGTTTCNC CANGNTCGAA AAGGCAAAGG TTAAAGAAAT TGCCANTAAA CTGGGTTAAA TGAANACTNT
     TTCANACCCC AANCAAAGNC GTNCNAGCTT TTCCGTTTCC AATTTCTTTA ACGGTNATTT GACCCAATTT ACTTNTGANA
.........................................................................................
401  TGAGTTTTTT NTANATAAAA ATAATTGAAA TAANANAAAT TTTCCTTNCC CACAAAAAAA AAAAAGAGNA AANNTAANCA
     ACTCAAAAAA NATNTATTTT TATTAACTTT ATTNTNTTTA AAAGGAANGG GTGTTTTTTT TTTTTCTCNT TTNNATTNGT
.........................................................................................
 +1                                                               X R L L R A C
481  NAAATAANTA NAAAAAAAGN GAAAAAGGAN AAATCAAAAA NAAAAAAGCG GCCGNTGAAT TNTAGACTCC TACGAGCATG
     NTTTATTNAT NTTTTTTTCN CTTTTTCCTN TTTAGTTTTT NTTTTTTCGC CGGCNACTTA ANATCTGAGG ATGCTCGTAC
.........................................................................................
 +1   V  R  P  H  S  L  S  R  D  R  V  V  R  V  P  R  A  P  I  V  S  *  Y  F  G  V
561  TGTGCGCCCT CACAGCCTGT CCCGGGACAG AGTGGTCCGC GTCCCCCGAG CCCCCATTGT GTCTTGATAC TTTGGGGTGC
     ACACGCGGGA GTGTCGGACA GGGCCCTGTC TCACCAGGCG CAGGGGGCTC GGGGGTAACA CAGAACTATG AAACCCCACG
.........................................................................................
 +1 H  M  A  I  D  L  Y  C  G  L  A  C  L  W  G  K  H  E  P  R  I  Y  T  D  W  A  N
641  ACATGGCTAT AGATCTTTAC TGCGGTTTGG CTTGTCTGTG GGGAAAACAT GAGCCCCGAA TCTACACAGA CTGGGCCAAT
     TGTACCGATA TCTAGAAATG ACGCCAAACC GAACAGACAC CCCTTTTGTA CTCGGGGCTT AGATGTGTCT GACCCGGTTA
.........................................................................................
 +1  H  Y  L  A  K  S  G  H  K  R  I  I  K  D  F  Q  Q  E  S  R  I  L  Q  I  S  I  T
721  CATTACCTAG CCAAATCCGG CCACAAGCGT ATCATCAAGG ATTTCCAGCA AGAAAGCCGA ATTCTGCAGA TATCCATCAC
     GTAATGGATC GGTTTAGGCC GGTGTTCGCA TAGTAGTTCC TAAAGGTCGT TCTTTCGGCT TAAGACGTCT ATAGGTAGTG
.........................................................................................
 +1   X  A  A  A  R  A  C  X  *  R  P  I  R  P  I  V  S  C  I  Q
801  ANTGGCGGCC GCTCGAGCAT GCATNTAGAG GCCCATTCGC CCTATAGTGA GTTGTATACA AT
     TNACCGCCGG CGAGCTCGTA CGTANATCTC CGGGTAAGCG GGATATCACT CAACATATGT TA
.........................................................................................
```

Figure 12. Tblastn search of the EST division of Genbank with 680aa starting at the c-terminus of the alfa-actinin domain of Hu-UNC-53/2.

```
LOCUS       AA418158     610 bp    mRNA            EST       19-MAY-1997
DEFINITION  zv97d12.r1 Soares NhHMPu S1 Homo sapiens cDNA clone 767735 5'.
ACCESSION   AA418158
NID         g2079968
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1..610
                     /organism="Homo sapiens"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
                     /clone="767735"
                     /clone_lib="Soares NhHMPu S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
     mRNA            <1..>610
```

```
                    /clone="5D16"
                    /clone_lib="Zebrafish ICRFzfls"
                    /sex="mixed"
                    /tissue_type="pooled 26-somite embryos"
                    /lab_host="XL1-blue MRF"
     mRNA           complement(<1..>418)
BASE COUNT      108 a     87 c     78 g    145 t
ORIGIN
        1 tttacatttt ttgaggaaga tgctaatggt ctattctgat tcaatgattt atgctaagct
       61 aagctaaaat gctcctgtca aatcctgaga tcagctgaat gaattaaaaa tttggtaaaa
      121 ctcaactgtc taactctagg ggagttgtaa aatgggccta tttccctaaa aagtaatgtt
      181 actttaagag catgatggtc caccagtttc actgtctaaa ttttgttatt ccataagcta
      241 atcttctctg ggcattttga cgattttaac actaacctgt gggtaatctg cgtccccegt
      301 aaactggaca tggtttcttc cagattctgt ctcagatcag caatgttctt cactgtacgc
      361 atccgtctag tttctggatc ttctcctgag atctcctcca ggcactgttt ggcggtct
//
gb|AA495042|AA495042 fa05f06.s1 Zebrafish ICRFzfls Danio rerio cDNA
           clone 5D16 3'
           Length = 418

  Minus Strand HSPs:

 Score = 195 (87.9 bits), Expect = 9.9e-18, P = 9.9e-18
 Identities = 37/46 (80%), Positives = 42/46 (91%), Frame = -3

Query:   627 TGQPALEELTGEDPEARRLRTVKNIADLRQNLEETMSSLRGTQVTH 672
             T +   LEE++GEDPE RR+RTVKNIADLRQNLEETMSSLRGTQ+TH
Sbjct:   416 TAKQCLEEISGEDPETRRMRTVKNIADLRQNLEETMSSLRGTQITH 279

 MOUSE  2

LOCUS       AA208994       527 bp    mRNA            EST       18-FEB-1997
DEFINITION  mw75e12.r1 Soares mouse NML Mus musculus cDNA clone 676558 5'.
ACCESSION   AA208994
NID         g1807004
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```

```
                 Tel: 314 286 1800
                 Fax: 314 286 1810
                 Email: mouseest@watson.wustl.edu
                 This clone is available royalty-free through LLNL ; contact the
                 IMAGE Consortium (info@image.llnl.gov) for further information.
                 MGI:416262
                 Putative full length read
                 vector to vector length is 535
                 Seq primer: -28m13 rev2 ET from Amersham
                 High quality sequence stop: 478.
FEATURES                  Location/Qualifiers
     source               1..527
                          /organism="Mus musculus"
                          /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                          polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                          was primed with a Not I - oligo(dT) primer [5'
                          TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTT 3'];
                          double-stranded cDNA was ligated to Eco RI adaptors
                          (Pharmacia), digested with Not I and cloned into the Not I
                          and Eco RI sites of the modified pT7T3 vector. Library
                          constructed and normalized by Bento Soares and M.Fatima
                          Bonaldo."
                          /clone="676558"
                          /clone_lib="Soares mouse NML"
                          /tissue_type="Liver"
                          /lab_host="DH10B"
     mRNA                 <1..>527
BASE COUNT      151 a    139 c    136 g    101 t
ORIGIN

        1 tgtctctgga tgagaagagc cgaacaatga gtcggtcagg ctccttccgg gatgggtttg
       61 aggaagttca tggatcctcc ctgtccttgg tttccagcac atcctccatc tactccacgc
      121 cagaagaaaa atgccagtca gagattcgaa agctgaggcg agacgtggat gcctcccagg
      181 aaaaggtgtc tgcgctgact acccagctga ctgcaaatgc tcaccttgtg gcagccttcg
      241 agcagagtct gggaaacatg accatcaggc tacagagttt aactatgacc gctgagcaga
      301 aggattcaga actgaacgag ttaagaaaaa ccatcgagct gctgaagaaa cagaatgcag
      361 ctgcccaggc tgccattaat ggagtgatta acacgccaga gctcaactgc aaaggaaatg
      421 gcagtgccag gctacagacc tacgcatccg cagcaacact cctccgacag tgtctccagt
      481 atcaatagcg ccaccagcca ctcaagtgtg ggcagcaaca tagagag

gb|AA208994|AA208994 mw75e12.r1 Soares mouse NML Mus musculus cDNA
            clone 676558 5'
            Length = 527

  Plus Strand HSPs:

 Score = 541 (243.9 bits), Expect = 2.3e-76, Sum P(2) = 2.3e-76
 Identities = 110/143 (76%), Positives = 114/143 (79%), Frame = +3

Query:  1511 SLDEKSRTMSRSGSFRDGFEEVHGXXXXXXXXXXXXXXPEEKCQSEIRKLRRELDASQE 1570
             SLDEKSRTMSRSGSFRDGFEEVHG              PEEKCQSEIRKLRR++DASQE
Sbjct:     3 SLDEKSRTMSRSGSFRDGFEEVHGSSLSLVSSTSSIYSTPEEKCQSEIRKLRRDVDASQE 182

Query:  1571 KVSALTTQLTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIEXXXXXXXX 1630
```

```
                 KVSALTTQLTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIE
Sbjct:   183 KVSALTTQLTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIELLKKQNAA 362

Query:  1631 XXXXXXGVINTPELNCKGNGTAQ 1653
                   GVINTPELNCKGNG+A+
Sbjct:   363 AQAAINGVINTPELNCKGNGSAR 431

 Score = 116 (52.3 bits), Expect = 2.3e-76, Sum P(2) = 2.3e-76
 Identities = 24/25 (96%), Positives = 25/25 (100%), Frame = +1

Query:  1661 RQHSSDSVSSINSATSHSSVGSNIE 1685
             +QHSSDSVSSINSATSHSSVGSNIE
Sbjct:   451 QQHSSDSVSSINSATSHSSVGSNIE 525


LOCUS       AA049124      337 bp    mRNA            EST       09-SEP-1996
DEFINITION  mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 479167 5'.
ACCESSION   AA049124
NID         g1528794
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:289911
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1..337
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
```

```
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                    T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                    14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                    State Univ., from 2 ]; double-stranded cDNA was ligated to
                    Eco RI adaptors (Pharmacia), digested with Not I and
                    cloned into the Not I and Eco RI sites of the modified
                    pT7T3 vector. Library went through one round of
                    normalization, and was constructed by Bento Soares and
                    M.Fatima Bonaldo."
                    /clone="479167"
                    /clone_lib="Soares mouse embryo NbME13.5 14.5"
                    /sex="unknown"
                    /tissue_type="embryo"
                    /dev_stage="13.5-14.5dpc total fetus"
                    /lab_host="DH10B"
     mRNA           <1..>337
BASE COUNT       80 a    101 c     97 g     59 t
ORIGIN
        1 catcctctgt gggcaccgag gtcaccgaga cccctgctca ttcagtcccc cacactagac
       61 tgttccaagc caatgaagag gaggagccag agaagaagga ggtatcagaa ctgcgctctg
      121 aactatggga aaaagagatg aagctcacgg atatccggtt ggaggccctc aactctgccc
      181 accagctgga ccagcttcgg gagaccatgc acaatatgca gttggaggtg gacctgctga
      241 aagcagagaa tgaccggctg aaggttgccc ccgggccctc ctcaggctgc actccagggc
      301 aggtccctgg gtcatcggct ctgtcgtccc ctcgacg

gb|AA049124|AA049124 mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus
            musculus cDNA clone 479167 5'
            Length = 337

  Plus Strand HSPs:

 Score = 206 (92.9 bits), Expect = 3.9e-19, P = 3.9e-19
 Identities = 42/60 (70%), Positives = 51/60 (85%), Frame = +3

Query:  1760 DSEAETVMQLRNELRDKEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLK 1819
              + E + V +LR+EL +KEMKLTDIRLEAL+SAHQLDQLRE M+ MQ E++ LKAENDRLK
Sbjct:    84 EPEKKEVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLK 263
//


LOCUS       AA185349      348 bp    mRNA            EST       07-JAN-1997
DEFINITION  mu51c03.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
            642916 5'.
ACCESSION   AA185349
NID         g1769059
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 348)
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:394908
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 336.
FEATURES             Location/Qualifiers
     source          1..348
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATACTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library constructed and
                     normalized by Bento Soares and M.Fatima Bonaldo."
                     /clone="642916"
                     /clone_lib="Soares mouse lymph node NbMLN"
                     /sex="male"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
     mRNA            <1..>348
BASE COUNT       93 a     95 c     78 g     82 t
ORIGIN
        1 attcggcact gaggggatga ataatccacc aaattagtgt gtacatagga gttgctgggc
       61 ccccccccac tcttatctgc tgtagctagc ctctccctaa gcctcgcatc ttctctaaat
      121 ctatctctgc gttcttacca cttgttctgg ccaatagaac tccggatcaa gaggcagaat
      181 tcctcagata gcatctccag cctcaacagc atcaccagcc attccagcat cggcagcagc
      241 aaagatgctg atgccaagaa gaaaaagaag aagagttggg taagtaaagg cttggagata
      301 ggcctgtgct aggagtcact caccctgttg cagggaactg acccccttt
//


gb|AA185349|AA185349 mu51c03.r1 Soares mouse lymph node NbMLN Mus
            musculus cDNA clone 642916 5'
            Length = 348
```

```
Plus Strand HSPs:

Score = 154 (69.4 bits), Expect = 4.4e-12, P = 4.4e-12
Identities = 27/42 (64%), Positives = 40/42 (95%), Frame = +1

Query:  1656 DLRIRRQHSSDSVSSINSATSHSSVGSNIESDSKKKKRKNWL 1697
             +LRI+RQ+SSDS+SS+NS TSHSS+GS+ ++D+KKKK+K+W+
Sbjct:   157 ELRIKRQNSSDSISSLNSITSHSSIGSSKDADAKKKKKKSWV 282
```

FIGURE 12a

```
"SIM output with parameters:
substitution scores in BLOSUM62
O = 12, E = 4"

Sequence 1: hu1,              1702 residues
Sequence 2: hu2,              2350 residues

List of local alignments with score >= 100.0
----------
46.8% identity in 1726 residues overlap; Score: 2538.0; Gap frequency: 9.3%


hu1,          78 DPESQRKRTVQNVLDLRQNLEETMSSLRGSQVTHSSLEMTCYDS--DDANPRSVSSLSNR
hu2,         639 DPEARRLRTVKNIADLRQNLEETMSSLRGTQVTHSTLETTFDTNVTTEMSGRSILSLTGR
                 ***  * *** *  *************** ***** ** *           **  **  *

hu1,         136 SSPLSWRYGQSSPRLQAGDAPSVGGSCRSEGTPAWYMHGERAHYSHTMPMRSPSKLSHIS
hu2,         699 PTPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSAPLRRQLASRGSS
                   ***** ************** *               *   *    * *        *

hu1,         196 RLEL-VESLDSDEVDLKS------GYMSDSDLMGKTMTEDDDITTG--------------
hu2,         759 VCHVDVSDKAGDEMDLEGISMDAPGYMSDGDVLSKNI-RTDDITSGYMTDGGLGLYTRRL
                      *     ** **        ***** *   *     **** *

hu1,         235 ---------------------------WDESSSISSGLSDASDNLSSEEFNASSSLNSLP
hu2,         818 NRLPDGMAVVRETLQRNTSLGLGDADSWDDSSSVSSGISDTIDNLSTDDINTSSSISSYA
                                            ** *** *** **  ****    * ***  *

hu1,         268 STPTASRRNSTIVLRTDSEKRSLAESGLSWFSESEEKAPKKLEYDSGSLKMEPGTSKWRR
hu2,         878 NTPASSRKNLDV--QTDAEKHSQVERNSLWSGDDVKKSDGGS--DSG-IKMEPG-SKWRR
                  **  ** *      ** ** *  *    *      *       ***  ***** *****

hu1,         328 ERPESCDDSSKGGELKKPISLGHPGSLKKGKTPPVAVTSPITHTAQ--SALKVAGK---P
hu2,         932 NPSDVSDESDKSTSGKKNPVISQTGSWRRGMTAQVGITMPRTKPSAPAGALKTPGTGKTD
                       * * *    **       **   * *  *  * * *       ***  *

hu1,         383 EGKATDKGKLAVKNTGLQRSSSDAGRDRLSDAKKPPSGIARPSTSG--SFGYKKPP-PAT
hu2,         992 DAKVSEKGRLSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPTANANSFGFKKQSGSAA
                   *   ** *  *     ** *****      ** *    *  *    *** **    *

hu1,         440 GTATVMQTG-----GSATLSKIQKSSGIPVKPVNGRKTSLDVSNSAEPGFLAPGARSNIQ
hu2,        1052 GLAMITASGVTVTSRSATLGKIPKSSAL-VSRSAGRKSSMDGAQNQDDGYLALSSRTNLQ
                 * *     *      **** ** ***   *    *** * *       * **   * * *

hu1,         495 YRSLPRPAKSSSMSVTGGRGGPRPVSSSIDPSLLSTKQGGLTPSRLKEPTKVASGRTTPA
hu2,        1111 YRSLPRPSKSNSRNGAGNRSS----TSSID-SNISSKSAGLPVPKLREPSKTALGSSLPG
                 ******* ** *    * *       **** *  * *  **    * ** * * *   *

hu1,         555 PVNQTDREKEKAKAKAVALDSDNISLKSIGSPESTPKNQASHPTATKLAELPPTPLRATA
hu2,        1166 LVNQTDKEKGISSDNESVASCNSVKVNPAAQPVSSPAQTSLQPGAKYPDVASPTLRRLFG
                  ***** **                      * * *      * *       **  *
```

```
hu1,   615  KSFVKPPSLANLDKV-NSNSLDLPSSSDTTHAS--KVPDLHATSSASGGPL--------P
hu2,  1226  GKPTKQVPIATAENMKNSVVISNPHATMTQQGNLDSPSGSGVLSSGSSSPLYSKNVDLNQ
                 *    *     **    *    *                  ** *  **

hu1,   664  SCFTPSPAPILNINSASFSQGLELMSGFSVPKETRMYPKLSGLHRSMESLQMPMS---LP
hu2,  1286  SPLASSPSSAHSAPSNSLTWGTNASSSSAVSKDGLGFQSVSSLHTSCESIDISLSSGGVP
            *    **       * *   *    *   * *        * ** * **     *   *

hu1,   721  SAFPSSTPVPTPPAPPAAP-TEEETEELTWSGSPRAGQLDSNQ---------------RD
hu2,  1346  SHNSSTGLIASSKDDSLTPFVRTNSVKTTLSESPLSSPAASPKFCRSTLPRKQDSDPHLD
            *   *             *         * * **     *                  *

hu1,   765  RNTLPKKGLRY----QLQSQEETKERRHSHTIGGLPESDDQSELPSPPALPMSLSAKGQL
hu2,  1406  RNTLPKKGLRYTPTSQLRTQEDAKEWLRSHSAGGLQDTAANSPFSSGSSVTSPSGTRFNF
            ***********    **  **  **   **  ***          *   *

hu1,   821  TNIVSPTAAT------TPRITRSNSIPTHEAAFELYSGSQM-GSTLSLAERPKGMIRSGSF
hu2,  1466  SQLASPTTVTQMSLSNPTMLRTHSLSNADGQYDPYTDSRFRNSSMSLDEKSRTMSRSGSF
                ***  *        *   *  *          *  *    *  ** *    * *****

hu1,   875  RDPTDDVHGSVLSLASSASSTYSSAEERMQSEQIRKLRRELESSQEKVATLTSQLSANAN
hu2,  1526  RDGFEEVHGSSLSLVSSTLSVYSTPEEKCQSE-IRKLRRELDASQEKVSALTTQLTANAH
            **    **** *** **  * **  **  *** ********  *****  ** ** ***

hu1,   935  LVAAFEQSLVNMTSRLRHLAETAEEKDTELLDLRETIDFLKKKNSEAQAVIQGALNASET
hu2,  1585  LVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIELLKKQNAAAQAAINGVINTPEL
            ********* *** **  *  *** ** **  ** **  *** *  *** * *  *  *

hu1,   995  TPK--------ELRIKRQNSSDSISSLNSITSHSSIGSSKDADAKKKKKKSWVYELRSSF
hu2,  1645  NCKGNGTAQSADLRIRRQHSSDSVSSINSATSHSSVGSNIESDSKKKKRKNW---LRSSF
              *          *** ** **** ** ** ***** **     * **** * *   *****

hu1,  1047  NKAFSIKKGPKSASSYSDIEEIATPDSSAPSSPKLQHGSTETASPSIKSSTLSSVGTDVT
hu2,  1702  KQAFGKKKSPKSASSHSDIEE--TTDSSLPSSPKLPHNGSTGSTPLLRNSHSNSL-----
              **  ** ****** *****  * *** ****** *        *     *    *

hu1,  1107  EGPAHPAPHTRLFHANEEEEPEKKEVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRET
hu2,  1755  --------------ISECMDSEAETVMQLRNELRDKEMKLTDIRLEALSSAHQLDQLREA
                            *    *    *  **  **  ************* **********

hu1,  1167  MHNMQLEVDLLKAENDRLKVAPGPSSGSTPGQVPGSSALS-SPRRSLGLALTHSFGPSLA
hu2,  1801  MNRMQSEIEKLKAENDRLK---SESQGSGCSRAPSQVSISASPRQSMGLS-QHSLNLTES
            *  ** *   *********      * **     *     * *** * **   **

hu1,  1226  DTDLSPMDGISTCGPKEEVT--LRVVVRMPPQHIIKGDLKQQEFFLGCSKVSGKVDWKML
hu2,  1857  TSLDMLLDDTGECSARKEGGRHVKIVVSFQEEMKWKEDSRPHLFLIGCIGVSGKTKWDVL
                    *    *    *         **          * *     *  **  ****  *  *

hu1,  1284  DEAVFQVFKDYISKMDPASTLGLSTESIHGYSISHVKRVLDAEPPEMPPCRRGVNN---I
hu2,  1917  DGVVRRLFKEYIIHVDPVSQLGLNSDSVLGYSIGEIKRSNTSETPELLPCGYLVGENTTI
            *  *   ** **    ** * ***   *  ****    **     * **   **    *   *
```

```
hu1,      1341 SVSLKGLKEKCVDSLVFETLIPKPMMQHYISLLLKHRRLVLSGPSGTGKTYLTNRLAEYL
hu2,      1977 SVTVKGLAENSLDSLVFESLIPKPILQRYVSLLIEHRRIILSGPSGTGKTYLANRLSEYI
               **  *** *  ****** *****  * * ***  ***  ************ *** **

hu1,      1401 VERSGREVTEGIVSTFNMHQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEAGS
hu2,      2037 VLREGRELTDGVIATFNVDHKSSKELRQYLSNLADQCNSENNAVDMPLVIILDNLHHVSS
               * * *** * *   ***    * * *  ****** *   *    * **** ** *    *

hu1,      1461 ISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLR
hu2,      2097 LGEIFNGLLNCKYHKCPYIIGTMNQATSSTPNLQLHHNFRWVLCANHTEPVKGFLGRFLR
                *  ** * ************ **    ***  **  **     *  **  *** * **

hu1,      1521 RKLVESDSDINANKEELLRVLDWVPKLWYHLHTFLEKHSTSDFLIGPCFFLSCPIGIEDF
hu2,      2157 RKLMETEISGRVRNMELVKIIDWIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGS
               *** *         **    ** ** * * **  *** ** **  ***  ******

hu1,      1581 RTWFIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQQDQS--KLYH
hu2,      2217 RVWFTDLWNYSIIPYLLEAVREGLQLYGRRAPWEDPAKWVMDTYPWAASPQQHEWPPLLQ
               * ** **** ****** *     *    *  * ****  ** ** **     *      *

hu1,      1639 LPPPTVGPHSIASPPEDRTVKDSTPSSLDSDPLMAMLLKLQEAANY
hu2,      2277 LRPEDVGFDGYSMPREGSTSKQMPPSDAEGDPLMNMLMRLQEAANY
               * *  **      * * * *  **    **** **  *******
----------
WARNING: 49 local alignments have not been reported because of score < 100.0
```

```
  "SIM output with parameters:
  substitution scores in BLOSUM62
  O = 12, E = 4"

Sequence 1: Cel,              1583 residues
Sequence 2: hu2,              2350 residues

List of local alignments with score >= 54.0
----------
32.8% identity in 504 residues overlap; Score: 490.0; Gap frequency: 6.9%

Cel,      1058 VIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKLTNGPATR
hu2,      1766 VMQLRNELRDKEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLKSESQGS
               *  *  **      *    **  *   *    * ***  *     *    ** * * *

Cel,      1118 ASSRASIPVIYD---------DEHVYDAACSST------SASQSSKRSSGCNSIKVTVNV
hu2,      1826 GCSRAPSQVSISASPRQSMGLSQHSLNLTESTSLDMLLDDTGECSARKEGGRHVKIVVSF
                 ***   *              *      *              * *  *    *  *

Cel,      1163 DIAGEISSIVNPDKEIIVGYLAMSTSQSCWKDIDVSILGLFEVYLSRIDVEHQLGIDARD
```

```
hu2,      1886 QEEMKWKEDSRPHL-FLIGCIGVS-GKTKWDVLDGVVRRLFKEYIIHVDPVSQLGLNS-D
                                   *         *      *       *     *       **   *      *     ***     *

Cel,      1223 SILGYQIGELRRVIGDSTTMITSHPTDILTSSTTIRMFMHGAAQSRVDSLVLDMLLPKQM
hu2,      1943 SVLGYSIGEIKRSNTSETPELLPCGY-LVGENTTISVTVKGLAENSLDSLVFESLIPKPI
               * *** ***   *      *                      ***     * *     ****    * **

Cel,      1283 ILQLVKSILTERRLVLAGATGIGKSKLAKTLAAYVSIRTNQ--SEDSIVNISIPENNKEE
hu2,      2002 LQRYVSLLIEHRRIILSGPSGTGKTYLANRLSEYIVLREGRELTDGVIATFNVDHKSSKE
                     *      **  *  *   * **  **  *   *    *                  *               *

Cel,      1341 LLQVERRLEKILRSKESCI-----VILDNIPKNRIAFVVSVFANV-PLQNNEGPFVVCTV
hu2,      2062 LRQYLSNLADQCNSENNAVDMPLVIILDNL--HHVSSLGEIFNGLLNCKYHKCPYIIGTM
               * *     *     *           ****             *             *  *

Cel,      1395 NRY--QIPELQIHHNFKMSVMSNRLE---GFILRYLRRRAVEDEYRLTVQMPSELFKIID
hu2,      2120 NQATSSTPNLQLHHNFRWVLCANHTEPVKGFLGRFLRRKLMETEISGRVRN-MELVKIID
               *        * ** ****        *  *    **  * ***   *  *     *    ** ****

Cel,      1450 FFPIALQAVNNFIEKTNSVDVTVGPRACLNCPLTVDGSREWFIRLWNENFIPYLERVARD
hu2,      2179 WIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAVRE
                 *       * * *    * *** ***  * **  ***** **  ***   ****   *

Cel,      1510 GKKTFGRCTSFEDPTDIVSEKWPW
hu2,      2239 GLQLYGRRAPWEDPAKWVMDTYPW
               *    **   ***   *   **
----------
35.5% identity in 112 residues overlap; Score: 165.0; Gap frequency: 1.8%

Cel,        11 IYTDWANRHLSKGSLSKSIRDISNDFRDYRLVSQLINVIVPINEFSPAFTKRLAKITSNL
hu2,        11 IYTDWANHYLTKSGHKRLIKDLQQDVTDGVLLAQIIQVVA--NEKIEDINGCPKNRSQMI
               *******  * *       * *   *   *   *   *  *  *  *     **

Cel,        71 DGLETCLDYLKNLGLDCSKLTKTDIDSGNLGAVLQLLFLLSTYKQKLRQLKK
hu2,        69 ENIDACLNFLAAKGINIQGLSAEEIRNGNLKAILGLFFSLSRYKQQQQQPQK
                    **   *    *     *     *  *** * * * * * ** ***   *  *
----------
24.8% identity in 163 residues overlap; Score: 80.0; Gap frequency: 3.7%

Cel,       877 GSQLSLASTT--AYGSLNEKYEHAIRDMARDLECYKNTVDSLTKKQENYGALFDLFEQKL
hu2,      1534 GSSLSLVSSTLSVYSTPEEKCQSEIRKLRRELDASQEKVSALTTQLTANAHLVAAFEQSL
               ** *** * *   *     **    **   * *       *  **          *  *** *

Cel,       935 RKLTQHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASH
hu2,      1594 GNMTIRLQSLTMTAEQK---DSELNELRKTIELLKKQNAAAQAAINGVINTPELNCKGNG
                  *         *              **      *    * *            * *

Cel,       995 RSSMSSSSKSSKQEKISLSSFGK-NKKSWIRSSLSKFTKKKNK
hu2,      1651 TAQSADLRIRRQHSSDSVSSINSATSHSSVGSNIESDSKKKKR
                             * **         *    *      ***
----------
58.6% identity in 31 residues overlap; Score: 74.0; Gap frequency: 6.5%
```

```
                            F
Cel,          653 GYPDNFEDSSSLSSGISDNNELDDISTDDLS
hu2,          840 GDADSWDDSSSVSSGISDT--IDNLSTDDIN
                  *  *   **** ******    *  ****
----------
42.9% identity in 60 residues overlap; Score: 64.0; Gap frequency: 6.7%

Cel,          984 RQPSLESVASHRSSMSSSSKSSKQEKISLSSFGKNKKSWIRSSLSK-FTKKKNKNYDEAH
hu2,         1661 RQHSSDSVSSINSATSHSSVGS---NIESDSKKKKRKNWLRSSFKQAFGKKKSPKSASSH
                  ** *  ** *  *  * **  *    *   *  *  * * ***    * ***       *
----------
22.0% identity in 91 residues overlap; Score: 56.0; Gap frequency: 0.0%

Cel,          140 SKLPSPRVATSATASATNPNSNFPQMSTSRLQTPQSRISKIDSSKIGIKPKTSGLKPPSS
hu2,          177 SRLSGPTARVSAAGSEAKTRGGSTTANNRRSQSFNNYDKSKPVTSPPPPPSSHEKEPLAS
                  * *  *    **  *              * *                 *      *  *

Cel,          200 STTSSNNTNSFRPSSRSSGNNNVGSTISTSA
hu2,          237 SASSHPGMSDNAPASLESGSSSTPTNCSTSS
                  *  *        * *  **        ***
----------
WARNING: 44 local alignments have not been reported because of score < 54.0
```

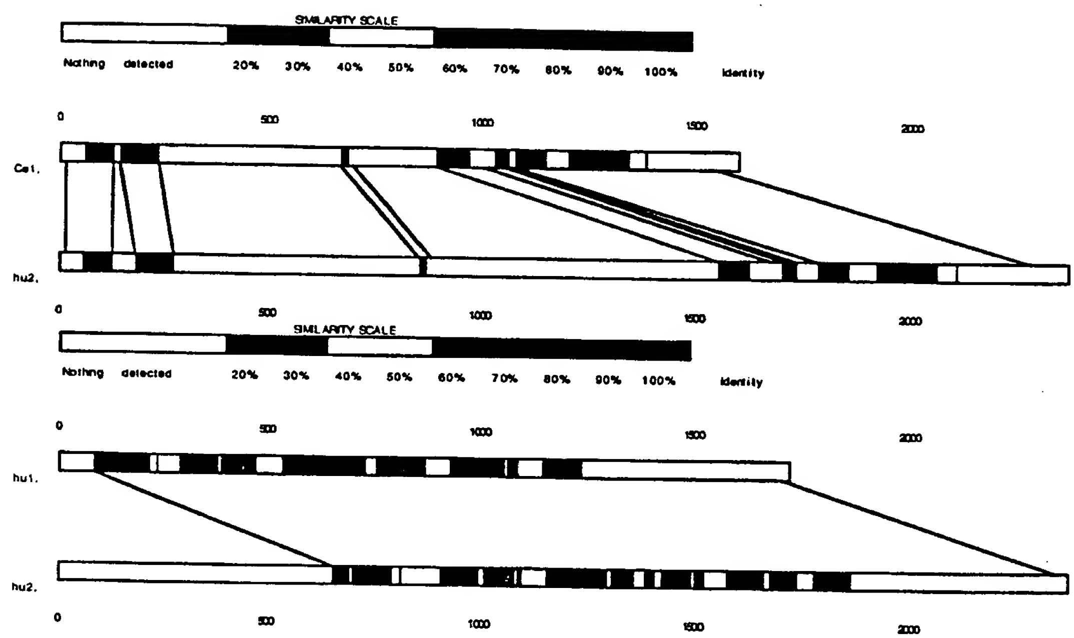

SIMILARITY SCALE
Nothing detected 20% 30% 40% 50% 60% 70% 80% 90% 100% Identity
0 500 1000 1500 2000
Ce1.
hu2.
0 500 1000 1500 2000
SIMILARITY SCALE
Nothing detected 20% 30% 40% 50% 60% 70% 80% 90% 100% Identity
0 500 1000 1500 2000
hu1.
hu2.
0 500 1000 1500 2000

Figure 12b

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence
Enzymes : 100 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

fig 13 6 pages

Page

```
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA  100
 T  D  R  E  I  S  R  S  P  M  V  D  S  Q  Y  N  L  L  .  C  R  I  V  K  P  V  S  A  P  C  L  C  V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACTGGCTGTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGC  200
 G  G  R  .  V  V  R  E  Q  N  L  S  Y  N  K  A  R  L  D  R  Q  L  H  E  E  S  A  .  G  .  A  F  C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGGTCTATATGCGCAACTGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT  300
A  A  S  R  C  T  G  Q  I  Y  A  L  T  L  I  I  D  .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCA  400
 G  V  P  R  Y  I  T  Y  G  K  V  P  A  W  L  T  A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG  500
N  A  N  R  D  F  P  L  T  S  M  G  G  L  F  T  V  N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A

CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT  600
P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V  H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTT  700
 R  Y  Y  H  G  D  A  V  L  A  V  H  Q  W  A  W  I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q

TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG
ACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTC  800
 W  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .  Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG  900
                                                                   T7 promotor priming site
G  L  Y  K  Q  S  S  L  A  N  .  R  T  H  C  L  L  A  Y  R  N  .  Y  D  S  L  .  G  D  P  S  W  L  A

GTTTAAACTTAAGCTTACCATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG  1000
                              ProBond binding domain
 F  K  L  K  L  T  M  G  G  S  H  H  H  H  H  H  G  M  A  S  M  T  G  G  Q  Q  M  G  R  D  L  Y  D
```

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GATGACGATAAGGTACCTAGGATCCATATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACA
CTACTGCTATTCCATGGATCCTAGGTATACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCTCTTTACGCAGCTGT 100

pCB201 insert = U4

U4 ORF

D D D K V P R I H M P P C R R G V N N I S V S L K G L K E K C V D

GCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCAC
CGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTG 200

pCB201 insert = U4

U4 ORF

S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S G P S G T

GGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAG
CCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTC 300

pCB201 insert = U4

U4 ORF

G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T F N M H Q

CAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATG
GTCAGAACGTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTAC 400

pCB201 insert = U4

U4 ORF

D S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L V I L L D

ACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGT
TGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACA 500

pCB201 insert = U4

U4 ORF

D - S E A G S I S E L V N G A L T C K Y H K C P Y I I G T T N Q P V

AAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGG
TTTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCCTACAACTGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCC 600

pCB201 insert = U4

U4 ORF

K M T P N H G L H L S F R M L T F S N N V E P A N G F L V R Y L F

```
BASE COUNT      173 a    168 c    141 g    128 t
ORIGIN
        1 gggccctcta gggtgcctgc tgcaggaagc acagcatagg tccagggagc ctctaattta
       61 aataggagaa gtcagagctt taacagcatt gacaaaaaca agcctccaaa ttatgcaaat
      121 ggaaacgaaa aagattcctc caaaggacct caatcgtctt caggtgtaaa tggtaacgtg
      181 cagcctccca gtactgctgg gcagcctcct gcctctgcca tcccttctcc aagtgccagc
      241 aagccctggc gcacgaagtc catgaatgtc aaacacagtg ccacctccac catgttgact
      301 gtaaagcagt caagtacagc cacctccccc acaccatctt cagacagact gaaggcaacc
      361 tgtctcagaa ggggtcaaaa ctgctccctc aggacagaaa tccatgcttg agaaattcaa
      421 gctagtcaat gcccggactg ctttacgccc cccgcagcct cccagttcag gacctagtga
      481 tggtgggaag gatgatgatg ccttttctga atctggtgaa atggaaggtt ttaacagtgg
      541 tctgaatagt ggtggctcaa caaatagcag tcccaaagtg tcacctaagt tggcccctcc
      601 aaaagctgga
```

```
LOCUS       AA495042      418 bp    mRNA            EST       27-JUN-1997
DEFINITION  fa05f06.s1 Zebrafish ICRFzfls Danio rerio cDNA clone 5D16 3'.
ACCESSION   AA495042
NID         g2225470
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Ostariophysi; Cypriniformes; Cyprinoidea; Cyprinidae; Rasborinae;
            Danio.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Clark,M., Lehrach,H., Johnson,S., Marra,M., Eddy,S., Hillier,L.,
            Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T.,
            Lacy,M., Le,N., Lennon,G., Martin,J., Moore,B., Schellenberg,K.,
            Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Steve Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Steve Johnson lab internal ID - P2_60 NOTE - For this library, the
            CLONE id field represents a position identifier on the original
            cDNA library preparation plate. cDNA Library Preparation: Matthew
            Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, and Max Planck Institut fuer Molekulare
            Genetik, Berlin Tel +49 30 84 13 1235
            Seq primer: -40m13 ET from Amersham
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1..418
                     /organism="Danio rerio"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3'], on
                     mRNA from pooled 26 somite zebrafish embryos;
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and  Max
                     Planck Institut fuer Molekulare Genetik, Berlin) and was
                     not biochemically normalised.  70,000 clones from this
                     library were arrayed on high density filters and
                     subsequently screened by oligonucleotide hybridization
                     fingerprinting to identify unique or minimally redundant
                     clones for more intensive analysis."
```

Tuesday, 18 November 1997 10:09 Page

fig 13 pCB201 (1 > 5082) Site and Sequence

```
AGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCT  1700
TCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAGCCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGA
                          pCB201 insert = U4
                          U4 ORF
 R  K  L  V  E  S  D  S  D  I  N  A  N  K  E  E  L  L  R  V  L  D  V  V  P  K  L  V  Y  H  L  H  T

TCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCT  1800
AGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGGCCTGGACCAAGTAACTGGA
                          pCB201 insert = U4
                          U4 ORF
F  L  E  K  H  S  T  S  D  F  L  I  G  P  C  F  F  L  S  C  P  I  G  I  E  D  F  R  T  W  F  I  D  L

GTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTC  1900
CACCTTGTTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTCACCTTACCCAG
                          pCB201 insert = U4
                          U4 ORF
  W  N  N  S  I  I  P  Y  L  Q  E  G  A  K  D  G  I  K  V  H  G  Q  K  A  A  W  E  D  P  V  E  W  V

CGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCG  2000
GCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGC
                          pCB201 insert = U4
                          U4 ORF
R  D  T  L  P  W  P  S  A  Q  Q  D  Q  S  K  L  Y  H  L  P  P  P  T  V  G  P  H  S  I  A  S  P  P

AGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTC  2100
TCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAG
                          pCB201 insert = U4
                          U4 ORF
E  D  R  T  V  K  D  S  T  P  S  S  L  D  S  D  P  L  M  A  M  L  L  K  L  Q  E  A  A  N  Y  I  E  S

TCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCCGGACAGCAGAACGCTGGCATCAGCTA  2200
AGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCTGTCGTCTTGCGACCGTAGTCGAT
                          pCB201 insert = U4
          U4 ORF
  P  D  R  E  T  I  L  D  P  N  L  Q  A  T  L  .  G  F  G  N  H  C  H  P  R  T  A  E  R  W  H  Q  L
```

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

TCTTAGCTCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTT 2300
AGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCCTCCTCTTGTCCTCCCTCCTCCTCTACTTTCTCCTCCCTGTCCAA

pCB201 insert = U4

S . L L L S P L L F Q S T G S P A P G G E Q E G G G D E R G G T G

CTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACT 2400
GAACCACGACATGGAAACTCTTGAAGGATCCTTCCTTACCACCCCACCGCAAACCCTTGAACACGGGGGATTTGTGTAAATGACCGGAGGAGATTACTGA

pCB201 insert = U4

S V C C T F E N F L G R N G G V A F G N L C P L N T F T G L L . . L

TTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGGTTCAGAAAACATCAAAACACTGC 2500
AACCCCTTTTCTACTAAGACCCAGAAAGGGAACTGAAGAACAAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCTTTTGTAGTTTTGTGACG

pCB201 insert = U4

W G K D D S G S F P . L L V S I T N S W A F W G G V Q K T S K H C

AGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGACAGTCTTGTGAGGGAGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCC 2600
TCGTCAAGGATTTACTAAGAGTGTTCGTTGGGACTCTCTCTGTCAGAACACTCCCTCTAGACCCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGTCTGGG

pCB201 insert = U4

S S S . M I L T S N P E R D S L V R E I W G R Q E A P Q I F S Q T

TTCCCAATTCCATCACCACTGCCAACACTCGTCCGGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTG 2700
AAGGGTTAAGGTAGTGGTGACGGTTGTGAGCAGGCCTTAAGACGTCTATAGGTCGTGTCACCGCCGGCGAGCTCAGATCTCCCGGGCAAATTTGGGCGAC

pCB201 insert = U4

L P N S I T T A N T R P E F C R Y P A Q W R P L E S R G P V . T R .

ATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC 2800
TAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCACGGAAGGAACTGGGACCTTCCACGGTGAGGGTGACAGGAAAGG

S A S T V P S S C Q P S V V C P S P V P S L T L E G A T P T V L S

TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA 2900
ATTATTTTACTCCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCCACCCCACCCCGTCCTGTCGTTCCCCCTCCTAACCCTTCTGT

. . N E E I A S H C L S R C H S I L G G G V G Q D S K G E D W E D

ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGC 3000
TATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCCTTTCTTGGTCGACCCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCG

N S R H A G D A V G S M A S E A E R T S W G S R G Y P H A P C S G A

ATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCC 3100
TAATTCGCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGG

L S A A G V V V T R S V T A T L A S A L A P A P F A F F P S F L A

ACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGG 3200
TGCAAGCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCGTAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCC

T F A G F P R Q A L N R G I P L G F R F S A L R H L D P K K L D .

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

```
GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
CACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACC 3300
G D G S R S G P S P . . T V F R P L T L E S T F F N S G L L F Q T G

AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTT
TTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACCCCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAA 3400
T T L N P I S V Y S F D L . G I L G I S A Y W L K N E L I . Q K F

AACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
TTGCGCTTAATTAAGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCA 3500
N A N . F C G M C V S . G V E S P Q A P Q A G R S M Q S M H L N .

CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCC
GTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGG 3600
S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P L T P

GCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCT
CGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGACGGAGA 3700
P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S A S

GAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGA
CTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTCGAGGGCCCTCGAACATATAGGTAAAAGCCTAGACTAGTTCTCT 3800
E L F Q K . . G G F F G G L G F C K K L P G A C I S I F G S D Q E

CAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACA
GTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGT 3900
T G . G S F R M I E Q D G L H A G S P A A W V E R L F G Y D W A Q Q

GACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTG
CTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGAC 4000
T I G C S D A A V F R L S A Q G R P V L F V K T D L S G A L N E L

CAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTAT
GTCCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATA 4100
Q D E A A R L S W L A T T G V P C A A V L D V V T E A G R D W L L

TGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGA
ACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACT 4200
L G E V P G Q D L L S S H L A P A E K V S I M A D A M R R L H T L D

TCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAA
AGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTT 4300
P A T C P F D H Q A K H R I E R A R T R M E A G L V D Q D D L D E

GAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGC
CTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACG 4400
E H Q G L A P A E L F A R L K A R M P D G E D L V V T H G D A C L
```

Tuesday, 18 November 1997 10:09

fig 13 pCB201 (1 > 5082) Site and Sequence

Page

CGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGA
GCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACT 4500
P N I M V E N G R F S G F I D C G R L G V A D R Y Q D I A L A T R D

TATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTT
ATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAA 4600
I A E E L G G E W A D R F L V L Y G I A A P D S Q R I A F Y R L L

GACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGA
CTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACT 4700
D E F F . A G L W G S K . P T K R R P T C H H E I S I P P P P S M

AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATT
TTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAA 4800
K G W A S E S F S G T P A G . S S S A G I S C W S S S P T P T C L L

GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
CGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGT 4900
Q L I M V T N K A I A S Q I S Q I K H F F H C I L V V V C P N S S

ATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAT
TACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTA 5000
M Y L I M S V Y R R P L A R A W R N H G H S C F L C E I V I R S Q

TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
AGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAAC 5082
F H T T Y E P E A . S V K P G V P N E . A N S H . L R W

FIG. 14.

```
    v360      v370      v380      v390      v400      v410      v420      v430      v440
SSVGTDVTEGPAHPAPHTRLFHANEEEKPEKKEVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLKVAP
SSVGT:VTE.PAH:.PHTRLF:ANEEEKPEKKEVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLKVAP
SSVGTEVTETPAHSVPHTRLFQANEEEKPEKKEVSELRSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLKVAP
        ^10       ^20       ^30       ^40       ^50       ^60       ^70       ^80       ^90

    v450      v460      v470      v480      v490      v500      v510      v520      v530
GPSSGSTPGQVPGSSALSSPRRSLGLALTHSFGPSLADTDLSPMDGISTCGPKEEVTLRVVVRMPPQHIIKGDLKQQEFFLGCSKVSGKV
GPSSG.TPGQVPGSSALSSPRRSLGLAL:H:F:PSL:DTDLSPMDGISTCG:KEEVTLRVVVRMPPQHIIKGDLKQQEFFLGCSKVSGKV
GPSSGCTPGQVPGSSALSSPRRSLGLALSHPFSPSLTDTDLSPMDGISTCGSKEEVTLRVVVRMPPQHIIKGDLKQQEFFLGCSKVSGKV
        ^100      ^110      ^120      ^130      ^140      ^150      ^160      ^170      ^180

    v540      v550      v560      v570      v580      v590      v600      v610      v620
DWKMLDEAVFQVFKDYISKMDPASTLGLSTESIHGYSISHVKRVLDAEPPEMPPCRRGVNNISVSLKGLKEKCVDSLVFETLIPKPMMQH
DWKMLDEAVFQVFKDYISKMDPASTLGLSTESIHGYS:SHVKRVLDAEPPEMPPCRRGVNNISV:LKGLKEKCVDSLVFETLIPKPMMQH
DWKMLDEAVFQVFKDYISKMDPASTLGLSTESIHGYSLSHVKRVLDAEPPEMPPCRRGVNNISVALKGLKEKCVDSLVFETLIPKPMMQH
        ^190      ^200      ^210      ^220      ^230      ^240      ^250      ^260      ^270

    v630      v640      v650      v660      v670      v680      v690      v700      v710
YISLLLKHRRLVLSGPSGTGKTYLTNRLAEYLVERSGREVTEGIVSTFNMHQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
YISLLLKHRRLVLSGPSGTGKTYLTNRLAEYLVERSGREVT:GIVSTFNMHQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
YISLLLKHRRLVLSGPSGTGKTYLTNRLAEYLVERSGREVTDGIVSTFNMHQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
        ^280      ^290      ^300      ^310      ^320      ^330      ^340      ^350      ^360

    v720      v730      v740      v750      v760      v770      v780      v790      v800
GSISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSDINANKEELLRVLDWVPKLW
GSISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSD:NANKEELLRVLDWVPKLW
GSISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSDVNANKEELLRVLDWVPKLW
        ^370      ^380      ^390      ^400      ^410      ^420      ^430      ^440      ^450

    v810      v820      v830      v840      v850      v860      v870      v880      v890
YHLETFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQQDQSKLYH
YHLETFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQQDQSKLYH
YHLETFLEKHSTSDFLIGPCFFLSCPIGIEDFRTWFIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQQDQSKLYH
        ^460      ^470      ^480      ^490      ^500      ^510      ^520      ^530      ^540

    v900      v910      v920      v930      v940      v950      v960
LPPPTVGPHSIASPPEDRTVKDSTPSSLDSDPLMAMLLKLQEAANYIESPDRETILDPNLQATL
LPPP:VGPHS.ASPPEDRTVKDSTP:SLDSDPLMAMLLKLQEAANYIESPDRETILDPNLQATL
LPPPSVGPHSTASPPEDRTVKDSTPNSLDSDPLMAMLLKLQEAANYIESPDRETILDPNLQATL
        ^550      ^560      ^570      ^580      ^590      ^600
```

FIG. 15.

SnaB I Nco I

TATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG 70

GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC 140

AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGA 210

CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACC 280

. L E N

Asp 718

Hind III Kpn I Bam HI

CACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTTGGTACCGAGCTCGG 350

P L L T G L S K L I R L T I G R P K L G T E L G

BstX I

Spe I Xma III EcoR I Pst I BspM II

ATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTGCAGATTATGCCATCAATTTCCGGATCTCAAGGA 420

S T S N G R Q C A G I L Q I M P S I S G S O G

M P S I S G S O G

ApaL I

ACTCTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAG 490

T L D N I D V I E L K Q E L K E R C S A L Y E

T L D N I D V I E L K Q E L K E R D S A L Y E

TCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTTGATGTTCTGAGGGAGACAGTGAACAAGTTGAAAAC 560

V R L D N L D R A R E V D V L R E T V N K L K T

V R L D N L D R A R E V D V L R E T V N K L K T

CGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCGC 630

E N K Q L K K E V C K L T N G P A T R A S S R

E N K Q L K K E V C K L T N G P A T R A S S R

GCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTC 700

A S I P V I Y D D E H V Y D A A C S S T S A S

A S I P V I Y D D E H V Y D A A C S S T S A S

FIG. 15 CONTINUED.

```
Asu II
AATCTTCGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAAT 770
C S S K R S S G C N S I K V T V N V D I A G E I
C S S K R S S G C N S I K V T V N V D I A G E I
Pvu I
Hpa I                           EcoR V
CAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCAGTCAGTCATGC 840
S S I V N P D K E I I V G Y L A M S T S Q S C
S S I V N P D K E I I V G Y L A M S T S Q S C
TGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATC 910
W K D I D V S I L G L F E V Y L S R I D V E H
W K D I D V S I L G L F E V Y L S R I D V E H
Cla I                                        Mlu I
AACTTGGAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGA 980
Q L G I D A R D S I L G Y Q I G E L R R V I G D
Q L G I D A R D S I L G Y Q I G E L R R V I G D
CTCCACAACCATGATAACCAGCCATCCAACTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCATG 1050
S T T M I T S H P T D I L T S S T T I R M F M
S T T M I T S H P T D I L T S S T T I R M F M
CACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCC 1120
H G A A Q S R V D S L V L D M L L P K Q M I L
H G A A Q S R V D S L V L D M L L P K Q M I L
AACTCGTCAAGTCAATTTTGACAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAA 1190
Q L V K S I L T E R R L V L A G A T G I G K S K
Q L V K S I L T E R R L V L A G A T G I G K S K
                         Asu II
ACTGGCGAAGACCCTGGCTGCTTATGTATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATC 1260
L A K T L A A Y V S I R T N Q S E D S I V N I
L A K T L A A Y V S I R T N Q S E D S I V N I
Bsm I                                        Bgl II
AGCATTCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGCA 1330
S I P E N N K E E L L Q V E R R L E K I L R S
S I P E N N K E E L L Q V E R R L E K I L R S
```

FIG. 15 CONTINUED.

Ava III
Nsi I
Xba I

AAGAATCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGC 1400

K E S C I V I L D N I P K N R I A F V V S V F A
K E S C I V I L D N I P K N R I A F V V S V F A

EcoR V

AAATGTCCCACTTCAAAACAACGAAGGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAG 1470

N V P L C N N E G P F V V C T V N R Y C I P E
N V P L C N N E G P F V V C T V N R Y C I P E

CTTCAAATTCACCACAATTTCAAAATGTCAGTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACC 1540

L Q I H H N F K M S V M S N R L E G F I L R Y
L Q I H H N F K M S V M S N R L E G F I L R Y

TCCGACGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCAT 1610

L R R R A V E D E Y R L T V D M P S E L F K I I
L R R R A V E D E Y R L T V D M P S E L F K I I

EcoR I

TGACTTCTTCCCAATAGCTCTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACA 1680

D F F P I A L Q A V N N F I E K T N S V D V T
D F F P I A L Q A V N N F I E K T N S V D V T

Bam HI

GTTGGTCCAAGAGCATGCTTGAACTGTCCTCTAACTGTCGATGGATCCCGTGAATGGTTCATTCGATTGT 1750

V G P R A C L N C P L T V D G S R E W F I R L
V G P R A C L N C P L T V D G S R E W F I R L

GGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCAC 1820

W N E N F I P Y L E R V A R D G K K T F G R C T
W N E N F I P Y L E R V A R D G K K T F G R C T

Bam HI
Tth I

TTCCTTCGAGGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAAT 1890

S F E D P T C I V S K K W P W F D G E N P E N
S F E D P T C I V S K K W P W F D G E N P E N

FIG. 15 CONTINUED.

```
                                          Tth I
GTGCTCAAACGTCTTCAACTCCAAGACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATC 1960
 V  L  K  R  L  Q  L  Q  D  L  V  P  S  P  A  N  S  S  R  Q  H  F  N
 V  L  K  R  L  Q  L  Q  D  L  V  P  S  P  A  N  S  S  R  Q  H  F  N
  Ava I
  Xho I
CCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTC 2030
 P  L  E  S  L  I  Q  L  H  A  T  K  H  Q  T  I  D  N  I  .
 P  L  E  S  L  I  Q  L  H  A  T  K  H  Q  T  I  D  N  I  .
                                                  Asp 718
                                                     Kpn I
TAATCTTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCCCC 2100
                                                        Ava I
                                                        Xma I
                                                          Sma I
CTTTTCCCCCCAATTTCCCAGAACCTCCTGTTCCCTTTGTTCCTAGTCCTCCCGGGTGCCGACGCCGAAG 2170
CGATTTAAAAACCTTTTTCTTTCCGAAACATTTCCCATTGCTCATTAATAGTCAAATTGAATAAACAGTG 2240
                                             Dra II
                                              Dra II
                                                Pss I
                                                 Apa I
                                                 Pss I
TATGTACTTAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGA 2310
     Bcl I
GCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTT 2380
CCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCT 2450
GAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAAT 2520
                                                              Pvu II
AGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGG 2590
GGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGC 2660
```

FIG. 15 CONTINUED.

Nae I

TACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC 2730

TTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC 2800

Dra III

CCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT 2870

GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCG 2940

GTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAAC 3010

AAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCA 3080

Ava III
Nsi I

GGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCC 3150

Ava III
Nsi I

CCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGC 3220

Nco I

CCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTA 3290

Stu I
Avr II

TGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTA 3360

Ava I
Xma I
Sma I
Bcl I

GGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGAT 3430

Xma III

CGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGG 3500

Nar I
Bbe I

CTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGC 3570

FIG. 15 CONTINUED.

Pst I

CCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTAT 3640

Bal I Fsp I Pvu II Tth I

CGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG 3710

GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCC 3780

ATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGA 3850

AACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA 3920

BssH II

GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTC 3990

Nco I

GTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCG 4060

Nae I Rsr II

ACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGA 4130

GCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATC 4200

Asu II

GCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGAC 4270

GCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTT 4340

Nae I

TTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACT 4410

TGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT 4480

Bsm I Sal I

TTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACC 4550

TCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC 4620

CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATT 4690

FIG. 15 CONTINUED.

Pvu II

AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGC 4760

CAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT 4830

CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 4900

GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTG 4970

CTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGC 5040

GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCC 5110

GACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCA 5180

ApaL I

CGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTC 5250

AGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC 5320

Alwn I

ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG 5390

TGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCT 5460

TCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG 5530

CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC 5600

BspH I

GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA 5670

TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTA 5740

CCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTC 5810

CCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG 5880

ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG 5950

TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCA 6020

FIG. 15 CONTINUED.

Fsp I

GTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGG 6090

CTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT 6160

Pvu I

TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCA 6230

Sca I

GCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCA 6300

AGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGC 6370

GCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATC 6440

ApaL I

TTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT 6510

TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG 6580

Ssp I BspH I

GAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG 6650

AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG 6720

Sal I Bgl II Sal I

TGCCACCTGACGTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCT 6790

Alwn I

GATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCA 6860

AAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTT 6930

Nru I Mlu I Spe I

TGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAA 7000

TCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCC 7070

CGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCC 7140

AATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA 7210

Nde I

GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC 7280

AGTACATGACCT 7292

FIG. 16.

```
              Sst I
TATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCAC  70

                                 Sst I
        Ppa I     Hind III        Bam HI Spe I    Xma III          EcoR I
TATAGGGAGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTG  140
                                                R  P  P  V  C  W  N  S

                                 Afl III
 Bgl II                          Mlu I
CAGATCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCA  210
A  D  L  G  Y  Q  I  G  E  L  R  R  V  I  G  D  S  T  T  M  I  T  S  H
                                                            M  I  T  S  H
TCCAACTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCATGCACGGTGCCGCACAGAGTCGCGTA  280
 P  T  D  I  L  T  S  S  T  T  I  R  M  F  M  H  G  A  A  Q  S  R  V
 P  T  D  I  L  T  S  S  T  T  I  R  M  F  M  H  G  A  A  Q  S  R  V
GACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACTCGTCAAGTCAATTTTGACAG  350
 D  S  L  V  L  D  M  L  L  P  K  Q  M  I  L  Q  L  V  K  S  I  L  T
 D  S  L  V  L  D  M  L  L  P  K  Q  M  I  L  Q  L  V  K  S  I  L  T
                                                    Bbv II
AGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTA  420
E  R  R  L  V  L  A  G  A  T  G  I  G  K  S  K  L  A  K  T  L  A  A  Y
E  R  R  L  V  L  A  G  A  T  G  I  G  K  S  K  L  A  K  T  L  A  A  Y
          Asu II                                    Bsm I
TGTATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAA  490
 V  S  I  R  T  N  Q  S  E  D  S  I  V  N  I  S  I  P  E  N  N  K  E
 V  S  I  R  T  N  Q  S  E  D  S  I  V  N  I  S  I  P  E  N  N  K  E
                                                         Ava III
  Xmn I                          Bgl II                  Nsi I     Xba I
GAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGCAAAGAATCATGCATCGTAATTCTAG  560
 E  L  L  Q  V  E  R  R  L  E  K  I  L  R  S  K  E  S  C  I  V  I  L
 E  L  L  Q  V  E  R  R  L  E  K  I  L  R  S  K  E  S  C  I  V  I  L
ATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGTCCCACTTCAAAACAACGA  630
D  N  I  P  ·  N  R  I  A  F  V  V  S  V  F  A  N  V  P  L  Q  N  N  E
D  N  I  P  ·  N  R  I  A  F  V  V  S  V  F  A  N  V  P  L  Q  N  N  E
                                 EcoR V
AGGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAA  700
 G  P  F  V  V  C  -  V  N  P  Y  Q  I  P  E  L  Q  I  H  H  N  F  K
 G  P  F  V  V  C  -  V  N  P  Y  Q  I  P  E  L  Q  I  H  H  N  F  K
```

FIG. 16 CONTINUED.

ATGTCAGTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGACGACGGGCGGTAGAGGATG 770
M S V M S N R L E G F I L R Y L R R R A V E D
M S V M S N R L E G F I L R Y L R R R A V E D

Sst I
AGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCTCTTCA 840
E Y R L T V Q M P S E L F K I I D F F P I A L Q
E Y R L T V Q M P S E L F K I I D F F P I A L Q

EcoR I
GGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAAC 910
A V N N F I E K T N S V D V T V G P R A C L N
A V N N F I E K T N S V D V T V G P R A C L N

Bam HI
TGTCCTCTAACTGTCGATGGATCCCGTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATT 980
C P L T V D G S R E W F I R L W N E N F I P Y
C P L T V D G S R E W F I R L W N E N F I P Y

Afl III Bam HI Tth I
TGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCACTTCCTTCGAGGATCCCACCGACAT 1050
L E R V A R D G K K T F G R C T S F E D P T D I
L E R V A R D G K K T F G R C T S F E D P T D I

Bbv II
CGTCTCTAAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAA 1120
V S K K V P W F D G E N P E N V L K R L Q L Q
V S K K V P W F D G E N P E N V L K R L Q L Q

Tth I Xho I
GACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCCTCGAGTCGTTGATCCAATTGC 1190
D L V P S P A N S S R Q H F N P L E S L I Q L
D L V P S P A N S S R Q H F N P L E S L I Q L

Bbv II
ATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATCTTCTCTCGCCTCTCCCCCG 1260
H A T K H Q T I D N I
H A T K H Q T I D N I
CTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCCCCCTTTTCCCCCCAATTTCCCAGAAC 1330

Xma I
Sma I
CTCCTGTTCCCTTTGTTCCTAGTCCTCCCGGGTGCCGACGCCGAAGCGATTTAAAAACCTTTTTCTTTCC 1400

Xmn I
GAAACATTTCCCATTGCTCATTAATAGTCAAATTGAATAAACAGTGTATGTACTTAAAAAAAAAAAAAAA 1470

FIG.16 CONTINUED.

```
                                                                     Sst I Bcl I
AAAAAAAAAAAAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCTGATCAGCCTCGACTGTG  1540
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTC  1610
CCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGG  1680
                                      Bbv II
GGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTG  1750
GGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCG  1820
GCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCC  1890
CGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGG  1960
GGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATG  2030
       Dra III
GTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAA  2100
TAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGG  2170
                                                                 Xmn I
ATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCT  2240
GTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCAT  2310
  Ava III
  Nsi I
GCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGC  2380
    Ava III
    Nsi I
ATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCA  2450
GTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGC  2520
                                                                 Xma I
                                             Stu I                  Sma I
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGA  2590
                    Bcl I
GCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATG  2660
                  Xma III
GATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAAT  2730
                                              Nar I
                                                 Bbe I
CGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGAC  2800
CTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTC  2870
     Fsp I              Tth I
CTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG  2940
GCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGG  3010
```

FIG. 16 CONTINUED.

```
CTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA  3080
CTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGA  3150
                            BssH II
ACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGC  3220
                                                                   Rsr II
TTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGG  3290
ACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCG  3360
CTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAG  3430
                             Asu II
TTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTT  3500
CGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATC  3570
CTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTT  3640
                                                        Bsm I
ACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTT  3710
                                       Sna I
GTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCA  3780
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCA  3850
TAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC  3920
                                                                   Sbo I
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTG  3990
CGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG  4060
                                                              Afl III
TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTG  4130
AGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC  4200
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGAT  4270
ACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT  4340
GTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTG  4410
                               ApaL I
TAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG  4480
                                                             Alwn I
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG  4550
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACT  4620
AGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT  4690
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA  4760
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGT  4830
                BspH I
TAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT  4900
```

FIG. 16 CONTINUED.

TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACC 4970
TATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA 5040
Ppa I
CGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT 5110
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT 5180
Fsp I
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTT 5250
GCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAAC 5320
Pvu I
GATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGT 5390
TGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC 5460
Sca I Sbo I
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC 5530
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT 5600
Xmn I
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCG 5670
EcoK ApaL I
ATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAA 5740
AAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTT 5810
Ssp I BspH I
CCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATT 5880
TAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGG 5950
Bgl II
Alwn I
GAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT 6020
CTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAG 6090
Nru I
GCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGC 6160
Afl III
Mlu I Spe I
CAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAT 6230
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC 6300
CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCA 6370
Nde I
ATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC 6440
CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC 6510
SnaB I
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA 6580
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG 6650
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCG 6720
GTAGGCGTGTACGGTGGGAGGTCTA 6745

FIG. 17.

GGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGC 70

Fsp I

CAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTAT 140

GGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG 210

Pvu I

GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGG 280

Sca I

CAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAAC 350

CAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC 420

GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA 490

ApaL I

TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTAC 560

TTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA 630

Ssp I BspH I

CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA 700

TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA 770

Sal I Bgl II Sal I

AGTGCCACCTGACGTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCT 840

Alwn I

CTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAG 910

CAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGT 980

Nru I Mlu I Spe I

TTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGT 1050

AATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG 1120

FIG. 17 CONTINUED.

```
CCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACG  1190

CCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATC  1260

      Nde I
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC  1330

                                                      SnaB I           Nco I
CCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG  1400

GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC  1470

ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAA  1540

CTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGG  1610

                                                                   Hind III
CTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTT  1680
 .  L  E  N  P  L  L  T  G  L  S  K  L  I  R  L  T  I  G  R  P  K  L

Asp 718      Bam HI                           BstX I
   Kpn I               Spe I    Xma III            EcoR I     Pst I
GGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTGCAGATCGCCGCTGCCACCT  1750
 G  T  E  L  G  S  T  S  N  G  R  Q  C  A  G  I  L  Q  I  A  A  A  T

                    Ava I
                    Xma I
                     Sma I                                          Pvu II
CAACCTTCGGACAACATTCGCTAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTGTCAGC  1820
 S  T  F  G  Q  H  S  L  R  S  P  G  Y  S  S  Y  S  P  H  L  S  V  S  A

                     Spe I
                        Sal I
TGATAAGGACACAATGTCTATGCACTCACAGACTAGTCGACGACCTTCTTCACAAAAACCAAGCTATTCA  1890
 D  Y  D  T  M  S  M  H  S  Q  T  S  R  R  P  S  S  Q  K  P  S  Y  S
       M  S  V  H  S  Q  T  S  R  R  P  S  S  Q  K  P  S  Y  S
GGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATCCACCGAGCACAGAATGGCGG  1960
 G  Q  F  H  S  L  D  R  K  C  H  L  Q  E  F  T  S  T  E  H  R  M  A
 G  Q  F  H  S  L  D  R  K  C  H  L  Q  E  F  T  S  T  E  H  R  M  A
```

FIG. 17 CONTINUED.

```
                Ava I                                         Bam HI
CTCTCTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCCTACTCGGCGCG  2030
A  L  L  S  P  R  R  V  P  N  S  M  S  K  Y  D  S  S  G  S  Y  S  A  R
A  L  L  S  P  R  R  V  P  N  S  M  S  K  Y  D  S  S  G  S  Y  S  A  R
   Ava I
TTCCCGAGGTGGAAGCTCTACTGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAA  2100
 S  R  G  G  S  S  T  G  I  Y  G  E  T  F  Q  L  H  R  L  S  D  E  K
 S  R  G  G  S  S  T  G  I  Y  G  E  T  F  Q  L  H  R  L  S  D  E  K
                                 Bam HI                          Nde I
TCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGCTAGCACGACAGCATATGGAT  2170
 S  P  A  H  S  A  K  S  E  M  G  S  Q  L  S  L  A  S  T  T  A  Y  G
 S  P  A  H  S  A  K  S  E  M  G  S  Q  L  S  L  A  S  T  T  A  Y  G
                                                                   Sal I
CTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGT  2240
S  L  N  E  K  Y  E  H  A  I  R  D  M  A  R  D  L  E  C  Y  K  N  T  V
S  L  N  E  K  Y  E  H  A  I  R  D  M  A  R  D  L  E  C  Y  K  N  T  V
                                                           Hind III
CGACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAAA  2310
 D  E  L  T  K  K  Q  E  N  Y  G  A  L  F  D  L  F  E  Q  K  L  R  K
 D  E  L  T  K  K  Q  E  N  Y  G  A  L  F  D  L  F  E  Q  K  L  R  K
              Cla I
CTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACGATTCAGGCAGGACATTGCTC  2380
 L  T  Q  H  I  C  R  S  N  L  K  P  E  E  A  I  R  F  R  Q  D  I  A
 L  T  Q  H  I  C  R  S  N  L  K  P  E  E  A  I  R  F  R  Q  D  I  A
ATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCT  2450
H  L  R  D  I  S  N  H  L  A  S  N  S  A  H  A  N  E  G  A  G  E  L  L
H  L  R  D  I  S  N  H  L  A  S  N  S  A  H  A  N  E  G  A  G  E  L  L
                                   Cla I    Cla I
TCGTCAACCATCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAG  2520
 R  C  P  S  L  E  S  V  A  S  H  R  S  S  M  S  S  S  S  K  S  E  K
 R  C  P  S  L  E  S  V  A  S  H  R  S  S  M  S  S  S  S  K  S  E  K
```

FIG. 17 CONTINUED.

```
                                                        Bam HI
CAGGAGAAGATCAGCTTGAGCTCGTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGT
______________________________________________________________________ 2590
 Q  E  K  I  S  L  S  S  F  G  K  N  K  K  S  W  I  R  S  S  L  S  K
 Q  E  K  I  S  L  S  S  F  G  K  N  K  K  S  W  I  R  S  S  L  S  K
                                          Nde I        BspM II
TCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAACTCT
______________________________________________________________________ 2660
F  T  K  K  K  N  K  N  Y  D  E  A  H  M  P  S  I  S  G  S  Q  G  T  L
F  T  K  K  K  N  K  N  Y  D  E  A  H  M  P  S  I  S  G  S  Q  G  T  L
                                                        ApaL I
TGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGC
______________________________________________________________________ 2730
  D  N  I  D  V  I  E  L  K  Q  E  L  K  E  R  D  S  A  L  Y  E  V  R
  D  N  I  D  V  I  E  L  K  Q  E  L  K  E  R  D  S  A  L  Y  E  V  R
CTTGACAATCTGGATCGTGCCCGCGAAGTTGATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGA
______________________________________________________________________ 2800
 L  D  N  L  D  R  A  R  E  V  D  V  L  R  E  T  V  N  K  L  K  T  E
 L  D  N  L  D  R  A  R  E  V  D  V  L  R  E  T  V  N  K  L  K  T  E
ACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCGCGCCTC
______________________________________________________________________ 2870
N  K  Q  L  K  K  E  V  D  K  L  T  N  G  P  A  T  R  A  S  S  R  A  S
N  K  Q  L  K  K  E  V  D  K  L  T  N  G  P  A  T  R  A  S  S  R  A  S
AATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCT
______________________________________________________________________ 2940
 I  P  V  I  Y  D  D  E  H  V  Y  D  A  A  C  S  S  T  S  A  S  Q  S
 I  P  V  I  Y  D  D  E  H  V  Y  D  A  A  C  S  S  T  S  A  S  Q  S
Asu II
TCGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTT
______________________________________________________________________ 3010
S  K  R  S  S  G  C  N  S  I  K  V  T  V  N  V  D  I  A  G  E  I  S
S  K  R  S  S  G  C  N  S  I  K  V  T  V  N  V  D  I  A  G  E  I  S
  Pvu I
     Hpa I                         EcoR V
CGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCAGTCAGTCATGCTGGAA
______________________________________________________________________ 3080
S  I  V  N  P  D  K  E  I  I  V  G  Y  L  A  M  S  T  S  Q  S  C  W  K
S  I  V  N  P  D  K  E  I  I  V  G  Y  L  A  M  S  T  S  Q  S  C  W  K
AGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATCAACTT
______________________________________________________________________ 3150
 D  I  D  V  S  I  L  G  L  F  E  V  Y  L  S  R  I  D  V  E  H  Q  L
 D  I  D  V  S  I  L  G  L  F  E  V  Y  L  S  R  I  D  V  E  H  Q  L
```

FIG. 17 CONTINUED.

```
Cla I                                                            Mlu I
GGAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGACTCCA 3220
 G  I  D  A  R  D  S  I  L  G  Y  Q  I  G  E  L  R  R  V  I  G  D  S
 G  I  D  A  R  D  S  I  L  G  Y  Q  I  G  E  L  R  R  V  I  G  D  S
CAACCATGATAACCAGCCATCCAACTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCATGCACGG 3290
 T  T  M  I  T  S  H  P  T  D  I  L  T  S  S  T  T  I  R  M  F  M  H  G
 T  T  M  I  T  S  H  P  T  D  I  L  T  S  S  T  T  I  R  M  F  M  H  G
TGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACTC 3360
 A  A  Q  S  R  V  D  S  L  V  L  D  M  L  L  P  K  Q  M  I  L  Q  L
 A  A  Q  S  R  V  D  S  L  V  L  D  M  L  L  P  K  Q  M  I  L  Q  L
GTCAAGTCAATTTTGACAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGG 3430
 V  K  S  I  L  T  E  R  R  L  V  L  A  G  A  T  G  I  G  K  S  K  L
 V  K  S  I  L  T  E  R  R  L  V  L  A  G  A  T  G  I  G  K  S  K  L
                                Asu II                               Bsm I
CGAAGACCCTGGCTGCTTATGTATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCAT 3500
 A  K  T  L  A  A  Y  V  S  I  R  T  N  Q  S  E  D  S  I  V  N  I  S  I
 A  K  T  L  A  A  Y  V  S  I  R  T  N  Q  S  E  D  S  I  V  N  I  S  I
                                                         Bgl II
TCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGCAAAGAA 3570
 P  E  N  N  K  E  E  L  L  Q  V  E  R  R  L  E  K  I  L  R  S  K  E
 P  E  N  N  K  E  E  L  L  Q  V  E  R  R  L  E  K  I  L  R  S  K  E
    Ava III
    Nsi I     Xba I
TCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATG 3640
 S  C  I  V  I  L  C  N  I  P  K  N  R  I  A  F  V  V  S  V  F  A  N
 S  C  I  V  I  L  C  N  I  P  K  N  R  I  A  F  V  V  S  V  F  A  N
                                                       EcoR V
TCCCACTTCAAAACAACGAAGGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCA 3710
 V  P  L  Q  N  N  E  G  P  F  V  V  C  T  V  N  R  Y  Q  I  P  E  L  Q
 V  P  L  Q  N  N  E  G  P  F  V  V  C  T  V  N  R  Y  Q  I  P  E  L  Q
AATTCACCACAATTTCAAAATGTCAGTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGA 3780
 I  H  H  N  F  K  M  S  V  M  S  N  R  L  E  G  F  I  L  R  Y  L  R
 I  H  H  N  F  K  M  S  V  M  S  N  R  L  E  G  F  I  L  R  Y  L  R
```

FIG. 17 CONTINUED.

CGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACT 3850

R R A V E D E Y R L T V Q M P S E L F K I I D
R R A V E D E Y R L T V Q M P S E L F K I I D

EcoR I

TCTTCCCAATAGCTCTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGG 3920

F F P I A L Q A V N N F I E K T N S V D V T V G
F F P I A L Q A V N N F I E K T N S V D V T V G

Bam HI

TCCAAGAGCATGCTTGAACTGTCCTCTAACTGTCGATGGATCCCGTGAATGGTTCATTCGATTGTGGAAT 3990

P R A C L N C P L T V D G S R E W F I R L W N
P R A C L N C P L T V D G S R E W F I R L W N

GAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCACTTCCT 4060

E N F I P Y L E R V A R D G K K T F G R C T S
E N F I P Y L E R V A R D G K K T F G R C T S

Bam HI Tth I

TCGAGGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCT 4130

F E D P T D I V S K K W P V F D G E N P E N V L
F E D P T D I V S K K W P V F D G E N P E N V L

Tth I Ava I Xho I

CAAACGTCTTCAACTCCAAGACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCCTC 4200

K R L Q L Q D L V P S P A N S S R Q H F N P L
K R L Q L Q D L V P S P A N S S R Q H F N P L

GAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATC 4270

E S L I Q L H A T K H Q T I D N I .
E S L I Q L H A T K H Q T I D N I .

Asp 718 Kpn I

TTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCCCCCTTTT 4340

Ava I Xma I Sma I

CCCCCCAATTTCCCAGAACCTCCTGTTCCCTTTGTTCCTAGTCCTCCCGGGTGCCGACGCCGAAGCGATT 4410

FIG. 17 CONTINUED.

TAAAAACCTTTTTCTTTCCGAAACATTTCCCATTGCTCATTAATAGTCAAATTGAATAAACAGTGTATGT 4480

Dra II
Dra II
Pss I
Apa I
Pss I

ACTTAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCG 4550

Bcl I

CTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG 4620

ACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA 4690

GGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG 4760

Pvu II

GCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTAT 4830

CCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACAC 4900

Nae I

TTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCC 4970

CCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAA 5040

Dra III

AAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT 5110

TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA 5180

TTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA 5250

TTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAG 5320

Ava III
Nsi I

GCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGC 5390

Ava III
Nsi I

AGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATC 5460

FIG. 17 CONTINUED.

Nco I

CCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAG 5530

Stu I
Avr II

AGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTT 5600

Ava I
Xma I
Sma I
Bcl I

TTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTT 5670

Xma III

CGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATG 5740

Nar I
Bbe I

ACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGT 5810

Pst I

TCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGG 5880

Fsp I
Bal I
Pvu II
Tth I

CTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGC 5950

TATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCAT 6020

GGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACAT 6090

CGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATC 6160

BssH II

AGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGT 6230

Nco I

GACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTG 6300

FIG. 17 CONTINUED.

```
Nae I          Rsr II
GGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTG 6370
GCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTT 6440
                                          Asu II
CTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCA 6510
ACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCG 6580
   Nae I
GGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTT 6650
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCAC 6720
Bsm I                                                       Sal I
TGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAG 6790
CTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC 6860
AACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 6930
                                              Pvu II
CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACG 7000
CGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTC 7070
GTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA 7140
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC 7210
GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC 7280
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC 7350
TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTG 7420
                                              ApaL I
TAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC 7490
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG 7560
```

FIG. 17 CONTINUED.

```
     AlwN I
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG  7630
GCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA  7700
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGC  7770
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA  7840
                                BspH I
GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT  7910
TTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT  7980
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGT  8050
CGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA  8120
CGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT  8163
```

Figure 18 : Phase contrast images of MCF-7 cells transfected with pCB201 (upper) compared to mock (control) transfected MCF-7 cells (bottom).

The control cells are spread out on the tissue culture plastic and exhibit few filopodia outgrowths. The transfected cells appear smaller because they are slightly rounded up and have multiple filopodia outgrowths (arrowhead) per cell.

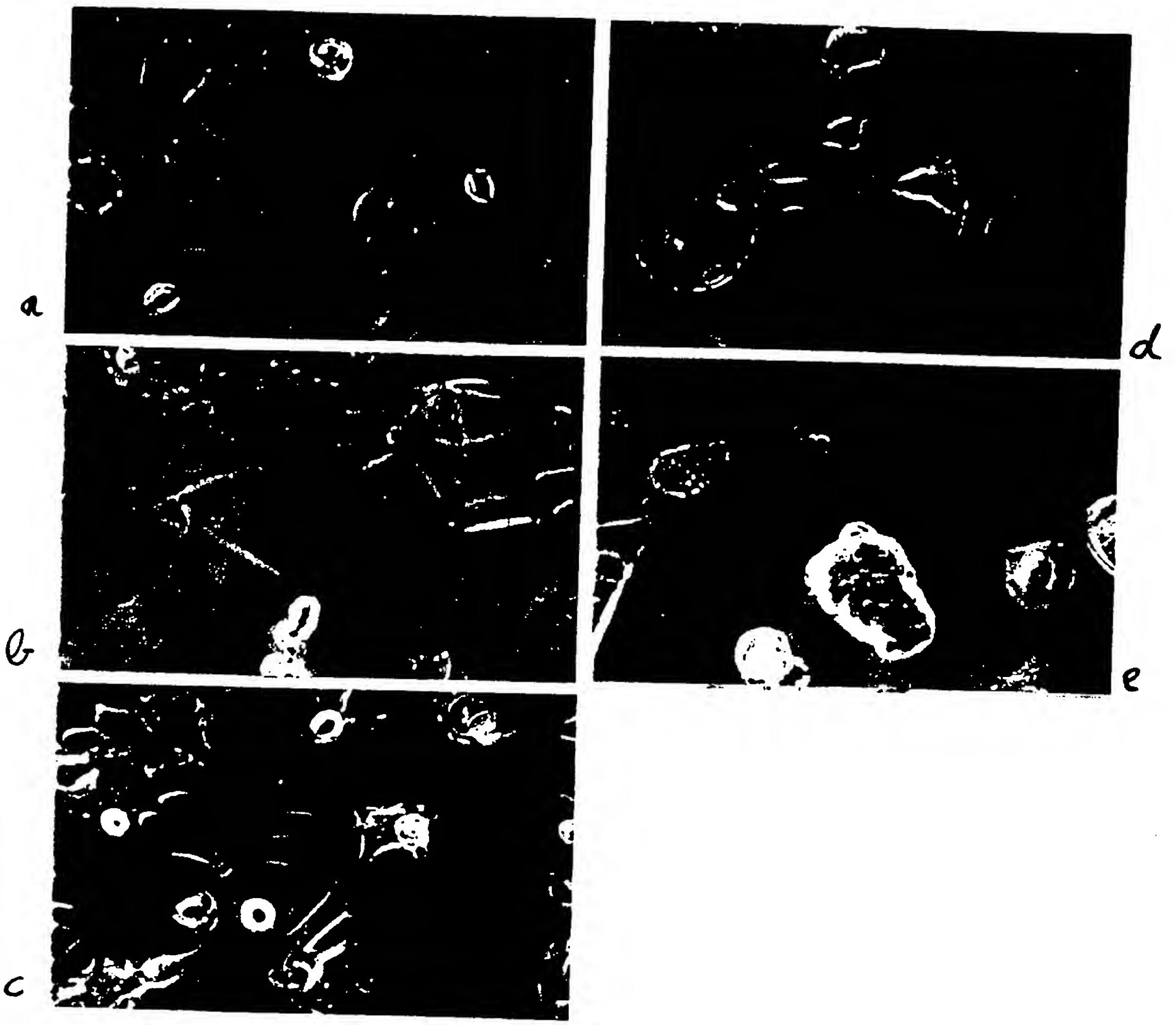


Figure 19 : Phase contrast image of MCF-7 cells, transfected with pcDNA3 (19a), pCDU4 (19b), pCDU3 (19c) pCDU2 (19d) and pTB72 (19e).

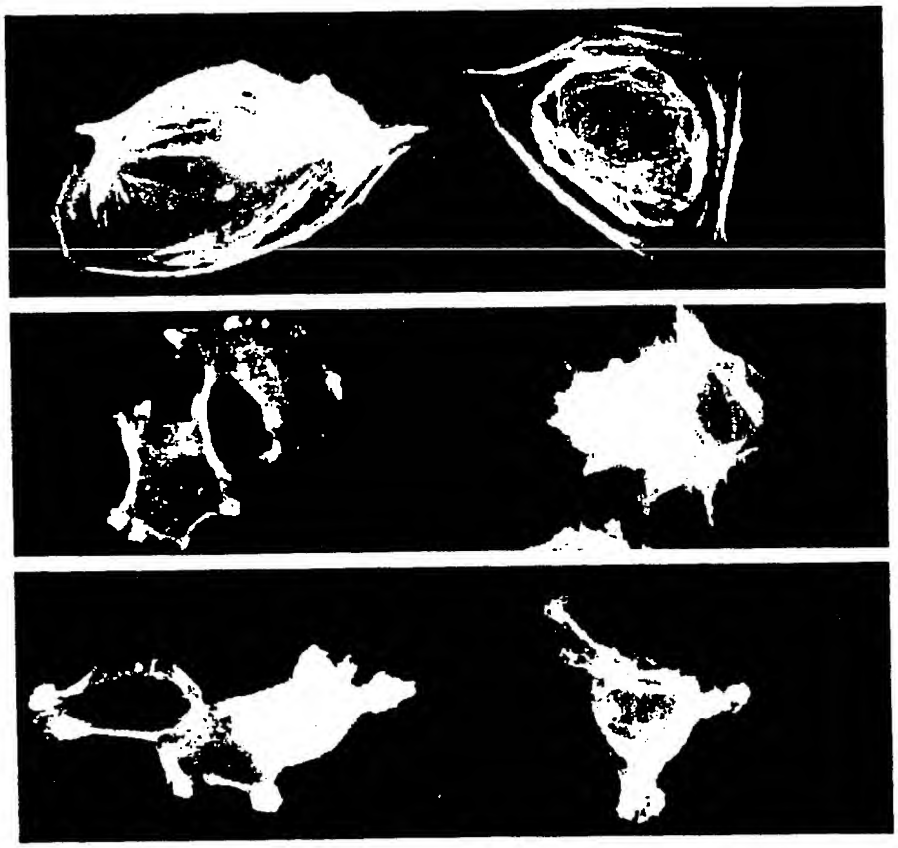


Figure 20 : F- actin pattern (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3.LacZ (top panel) and with pCB201 (middle and lower panel).

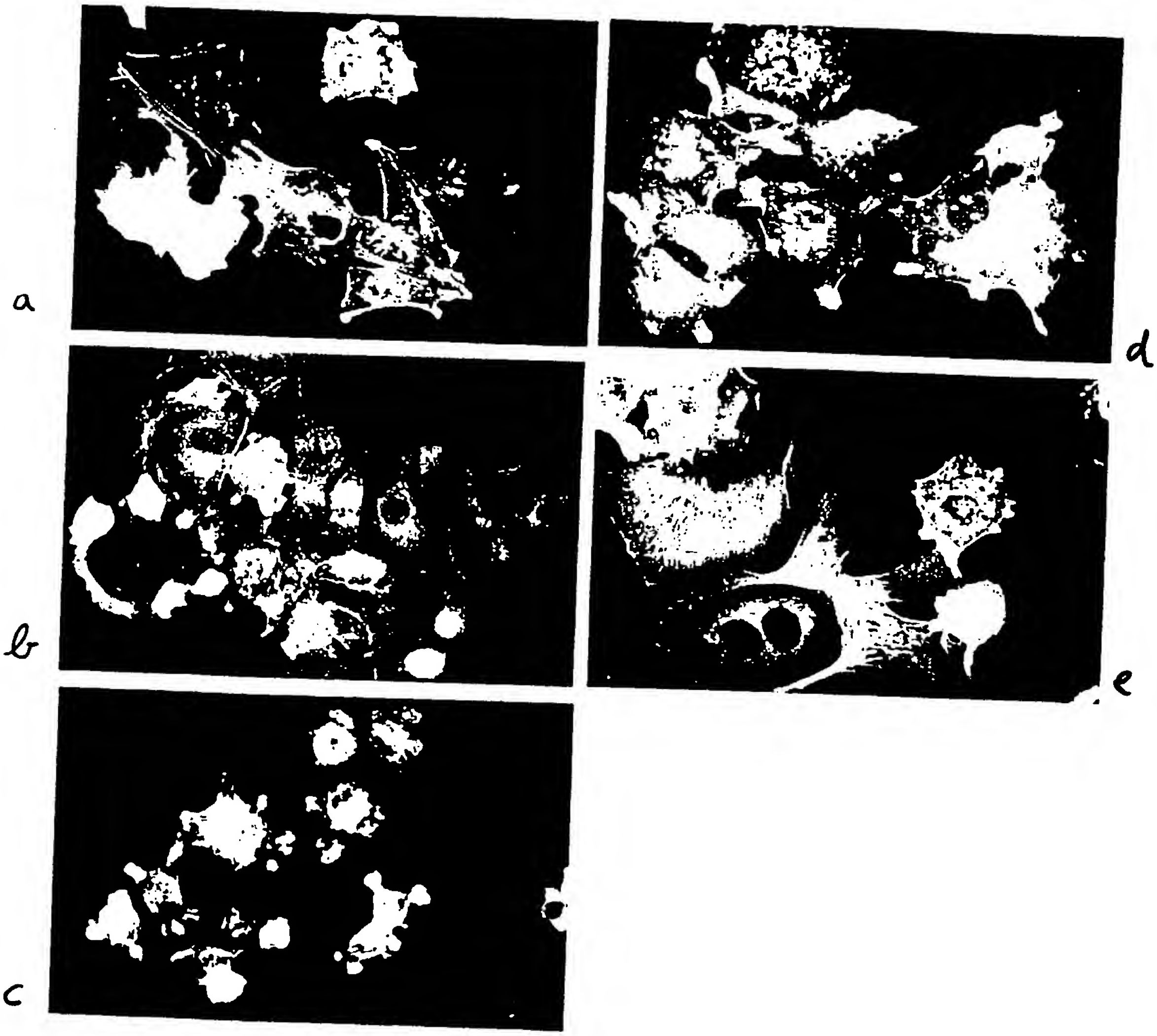


Figure 21 : F- actin pattern Phalloidin (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3 (21a), pCDU4 (21b), pCDU3 (21c) pCDU2 (21d) and pTB72 (21e)

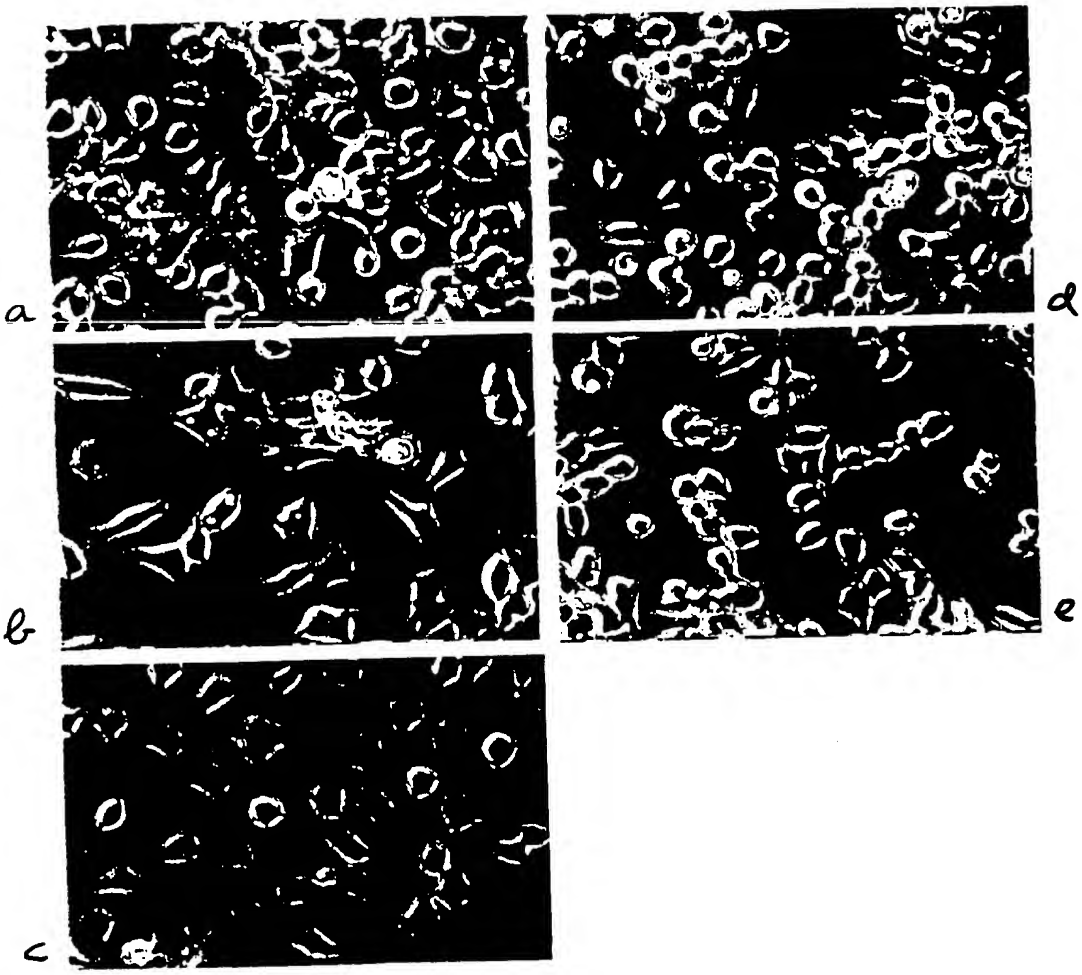


Figure 22 : Phase contrast image of N4 neuroblastoma cells, transfected with pcDNA3 (22a), pCDU4 (22b), pCDU3 (22c) pCDU2 (22d) and pTB72 (22e)

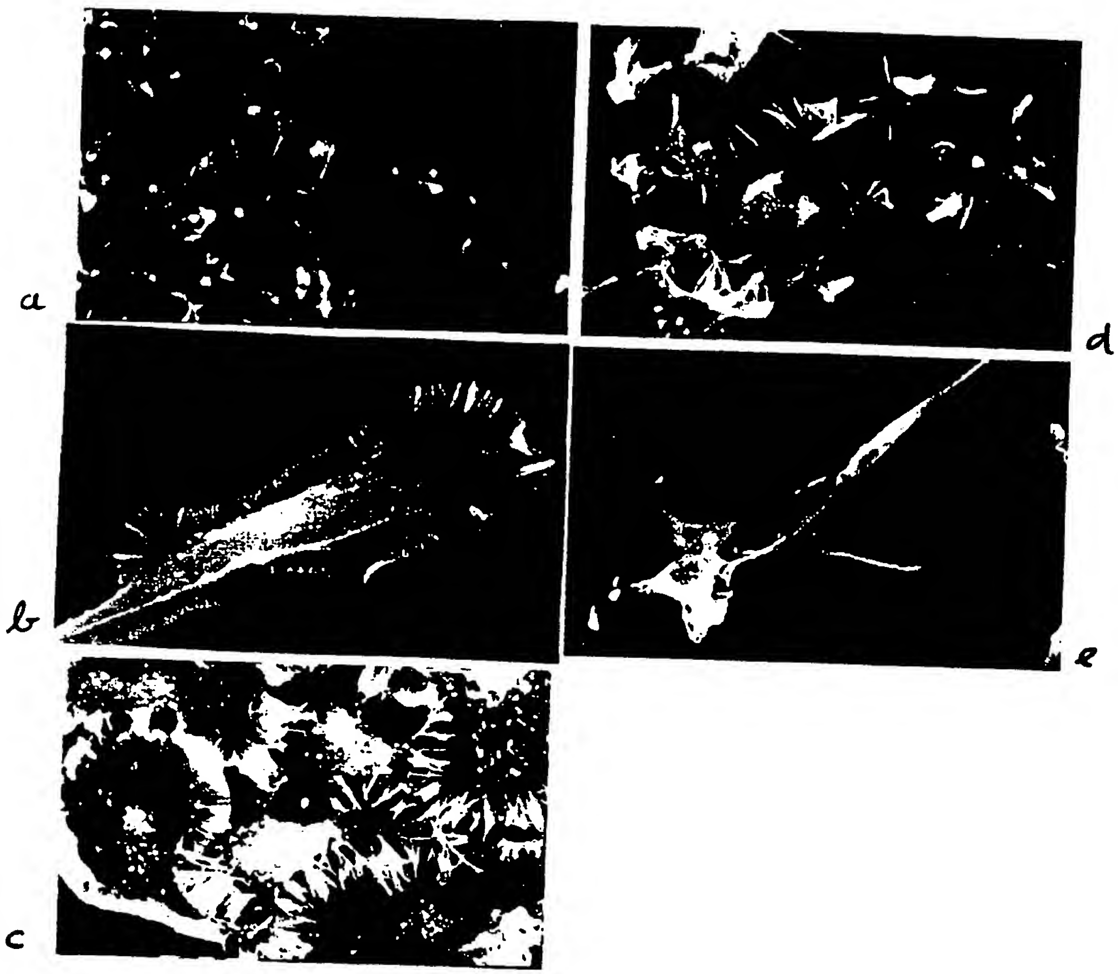


Figure 23 : F- actin pattern Phalloidin (visualized with TRITC-Phalloidin) of N4 neuroblastoma cells transfected with pcDNA3 (23a), pCDU4 (23b), pCDU3 (23c) pCDU2 (23d) and pTB72 (23e)

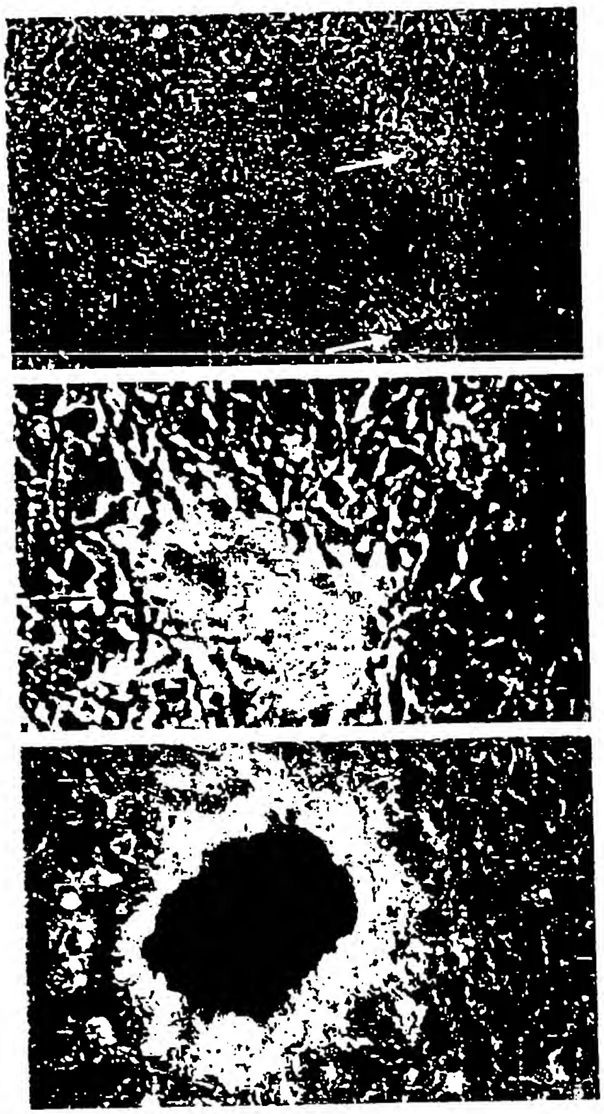


Figure 24 : Phase contrast images of small, medium sized and large foci induced in a monolayer of NIH-3T3 cells by transfection with pCB201.

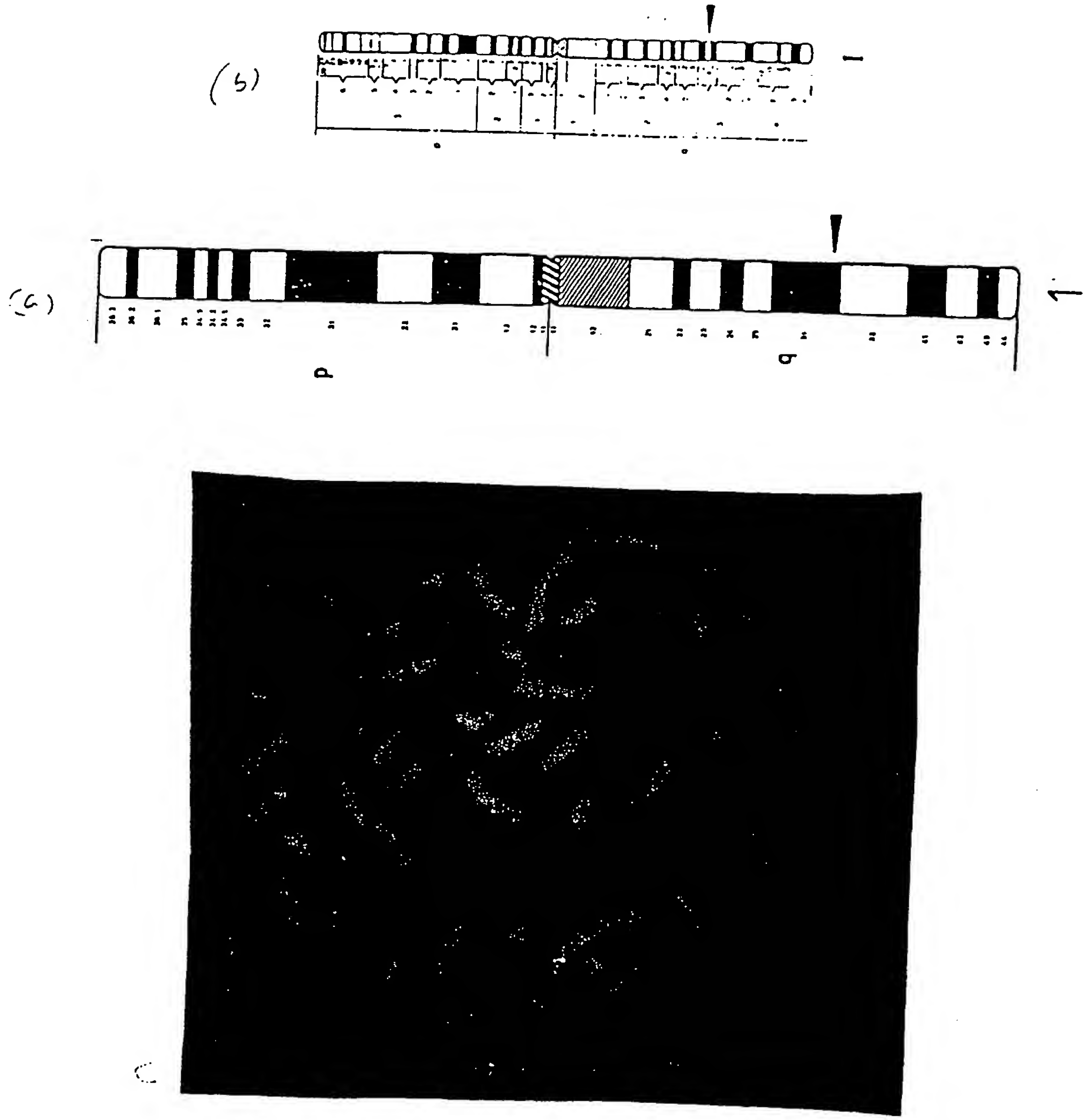
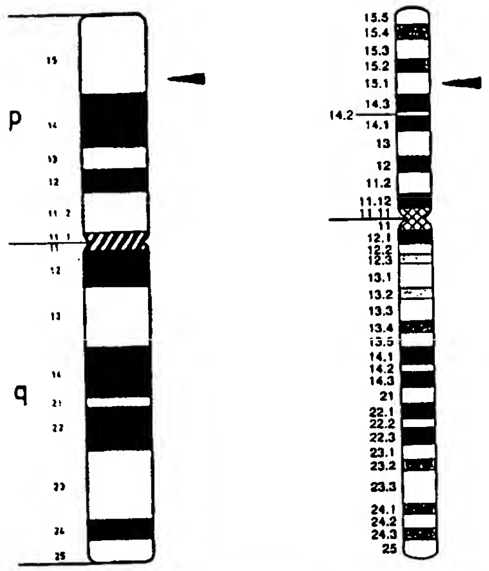


Figure 25a. b. c: Chromosomal localisation of hu-unc-53/1 by FISH

(d) (e)



11

(f)

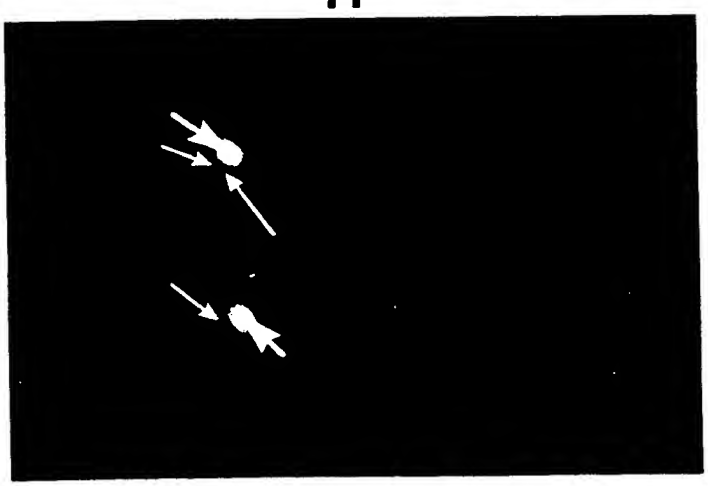
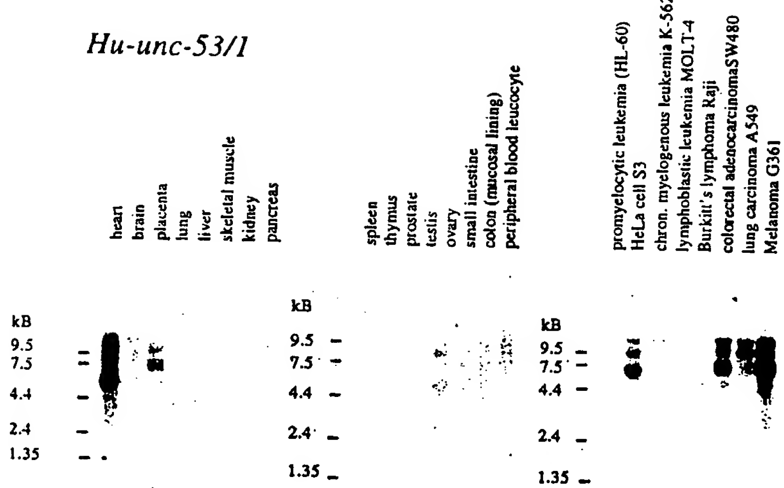


Figure 25d, e, f: Chromosomal localisation of hu-unc-53/2 by FISH



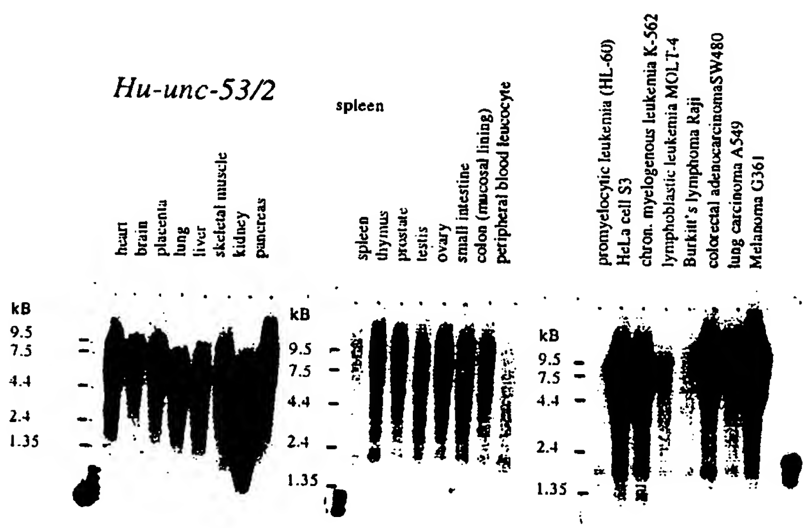


Figure 26 : Expression pattern of *Hu-unc-53/1* and *Hu-unc-53/2* in normal human tissues and cancer cell lines by Northern blotting.

donderdag, 27 november 1997 16:52 Page 1
fig 27 pNP3 Map (1 > 13621) Site and Sequence
Enzymes : All 146 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTAT 100

AAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAA 200

CCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAG 300

CCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG 400

GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGAT 500

CGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTG 600

TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGG 700

CTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGAGCGGGGGACGCGCAGGACACACAGTTTAGTTATTTATGAGGAAGAGGGAGAAAG 800

ATTATGCAAATGACAAAGAGAGGAGAATGGTGACCAGCCACCAGTGTCGGCAAGTTTCTTCTTGTCAAGTGGCTACTCGAACACATACTATCAATTGGTT 900

TTTGCTTGTGTGTGTTCCAAATATGGAGGCACACTGGTTTGAGGTTGAGGAGGCGACTTCAAAGACTTATACCTTGGAAACAAATACAAAAATCAAAAAA 1000

AGTCAACTAGGATTTCAAATAAAAAATTTCAAACATTTAGATTTTTTTATAAGTTTTTCTAATTTAGAAGTTCGCTTTCCAGATAAATTGCATCAAAAAT 1100

CTAGGCTTTCAATTTTTGAATTTTCTCGAAGTGATCAACTCAAATTTGGCTTTAATATTCAAAATTTGAATTTTTAAATTAAAAAAAAACCTTTTTTTCTT 1200

TCTTAAACTGTTTCTAGTAAAAAAGCTCCAAATTATTTTTGCATAATATTCCTTAGAAGTACATTTTTTATCAGCTCATCGGGAGCTTTGGCACCTATGG 1300

GCTCTGCTAGGCCCCGCCACAAAAACCTTCCGAGGAACTGTAATTTTTGTGGCGAGACCCACGGAACGTAGTCATGATGGGAGTCAAATGGAAGGCGCAT 1400

GGAATTCAAATTTGGCAGTTTAATCTTTAAAAAAATTGGAATTTCGAAAAATATGATTTCGGAAATTTTTGAAAAAAACGTTCTAAGGTTATAGTTGAGA 1500

TAATTTGAGATCCGAATTGACACACATTTTTGGCTAGAGTTTGAAAAAGTGGGTCCCGCCACGAAAACCTCCCTACAAAAGTTTTTATGAAACAAATTTT 1600

GAAATAAAACCAAGATATTGAACACAATGCTGAGCAACATGGTGCATACTAACAATAATTTGGCAAGCTACCAAAAACCAAAAATTCTTGACAAACACTT 1700

CACTATCACAAAACGAATTCAATTCCGCCAAAAAGATGGTAAGAGGAGAAAAAGAAACCCCGTTCGTTTGTTCTCGGCTTAATGAATTAATTATAGGAAC 1800

ACAAGGCAATCGGAAGGAAACCGGCGACGGAAGGGAAGGGAAGGAAGCAAGTCGCCAAAAATGGAAGCGGGAGGACAGCTGCATTAGAGAGAGAGGGGAG 1900

CTTGTGCGCCAAATCTAGGGTGAATTGAAAGAAGATTGAAACCCACAGGATCGGGGTGACATGAAAAAGAAATTGAGACGAAGAAGAATGATAGGTTCGC 2000

GTTATTTCATAACCTAAATTAAATGGACAGGACTTGAAAGGAAATTTTGAATTTAGGCAGGGCTCAAGCGGAACCTTTTTGAGCTATAGTATGCAGACTC 2100

ACAGAACTACAGAATTTCAAAATTCATGAAACTTTCGTAATACCAACTATCTTGCCCATCTCCACTGAAACATCTCAAATTTCATAACTAGTTTCAAATT 2200

CAGAGGCATAACAAAAAAGCAACAAAAGCAAAGAAATGCTCTTCGAAGCGCGCAAAACCGCATCGGTTCCATTATGAGATGCACCGATAGATCTTCGTTA 2300

GAAAGAGAGAGAGTAAATAAACATTTGGATATCCTCCTAGGAGACAACTTGCGCTAAACCAGATGTTGGTTCTAGTATTTTTGAAAAAATGCATTTGGCG 2400

GCAACTTCTTCCCAAAGTTTGGCCCAAAAAGTATTGGACCAAGCTTGGTACAGACTCTAGAAAGACTTGGGCCAGCAAGACTTGTTGAGCAAATTCTTGC 2500

CACTGTTTGAAAATTACGCGTGACCCAAACAAAATTTGGCTCAAGTTTTACCAAACTGCCGCCAAATGCCAAGTAGACTGAGGTCACCGGAATATTCCTG 2600

TAATTTGCAATCTAAGCTCCAAACTATTTCACGCGGTTCCTCGATGCTCATTCATAAGTCATAGCTCCTCCTCCTCCTCTCATCTTTAGGAACCAAAATT 2700

GCACTTCGCGCATGAATAGATGTCTGTGTAACAGGGACTCCCAGCAGAGAGAGAAAAAGAATGCAAAAGAAGAAGATGCTATCAAACGTCTTTCGGGGAT 2800

TCCTCGATCACCAAGTGATGCTCATTAATTTTGAGCCTTCTGGATTAATTGCTCCATGGCTCCCTGACCACCGAAACGAGCAGCGAATGTGTTCATATGT 2900

GTGTCAAGGGACTACAACTACTACTACCAAAACTCACTTGTCAGGATAGCCGTTCTGTGCTCGACGAGTCATGTCATCCTATTCTTTCAGCTTCTCCCTC 3000

ATTTTGGCGCACATCGCAACACCGGCGTCAGAGGTGTAGCCGCTGGATTCTGAAATATTTTGAGAGTTATGAGGATTCAACTGTCGAAGGTTTTTTTTCA 3100

AATTCTAACGCCAATGTGCTCAAAAATTACTGTCGAAATTAAAAATCTACAAGTGTTCATTCACCGAATTAGTGATTAACGTTCAAAATGTAGACAACAA 3200

GGAATAGGTAAGCATGTAGGCAAATGTAAGGTGTTGGTAGGTCCAATAAACAATAATTATTAGCTTTTGATTTATTCAGTGTTTTTTTTTAAATATTATT 3300

TACTGTTTATTGGCTTGGTAAATATCAAGCCTATCTGCCTACATCATACCTTTCTGCCTACACACCTGCCTATGCCTAAAACTTAATACATCTCAAGAGA 3400

ACTTTCACAATTCCAAAAATTAAAATACTATAGATTCAAAAATTATCAAAAGAAAAAGTTTTTGGTACGAGCTCCAACGTTGGTTAACTTTAAACGAAAC 3500

AGTTCAGAAACCAAAAAATCAAAAGTTTTCCTTTGACAAACATCATTTTTCTGGGATTTTTTAAAATTTATAATTTGAAGCTATTTTTCAACAAAAAAAA 3600

TCAAAGTTTCCAAAATTTACCTGATGTCTTATTCTTTCCCATCCTTCTCTCCAGCGAATGATTACCAGAAGTTTTTGTCGGCGGAGTCACCTGAGCCGAC 3700

GCATGAGCCACAATGGAGTCTTCAGACGCGGACGACTGCTCATATCCTGCAAAATGGTCATTTGTCATAATAATTGGAGGTGGGGGGTGCACTGACCAAA 3800

CGACTTGACAGGCGATGTGATTTTTCCTTGTTTTAGAAGAACATCGTACGATAAGGTATTTTGTGTGCGGTTTTTTCTATGATGTTTCAAAAAAAATTTA 3900

TCAGAAGTAAAAAAAACATAAGAAGTGTACACAGAACATTGTGTAGATTTTGGGATAGAATTATTATTAATGGGGGGAGGGAGCACAATTTTTTTCGACT 4000

TTTTATTACGAATTAATTCTATTTTTGAATCTACACAATTTCATTGTTAGAGGTGACTTAAAAGAAGCTAAAAAACAAAGAAATTAAGAAGAAGTTTTTT 4100

TGATAATAACAAAAATAGGGGGTGGGAGCACAGGGAGAAAGAGCATGTAGGGATGTTGAAGAGTAATTGGACTTAGAAGTCAGAGGCACGGGCGCGAGAT 4200

GCTGAAAATTTAAATATTACTCGTTAGTTAGTAGAACTCAGTTTAAACTTACTCTGATGACAGCGGCCCCTATTATTTTTGACACCAGACAAGTTGGTAA 4300

TGGTAGCGACCGGCGCTCAGTTGGAATTCTACGAATGCTATTTGTATAGTTCATCCATGCCATGTGTAATCCCAGCAGCTGTTACAAACTCAAGAAGGAC 4400

CATGTGGTCTCTCTTTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGCCATCGCCAATTGGAGTA 4500

TTTTGTTGATAATGGTCTGCTAGTTGAACGCTTCCATCTTCAATGTTGTGTCTAATTTTGAAGTTCTGAAAATTTAAATATAATCCGTTAGTTAGTAAGT 4600

CCAAGTTTAAACTTACAACTTTGATTCCATTCTTTTGTTTGTCTGCCATGATGTATACATTGTGTGAGTTATAGTTGTATTCCAATTTGTGTCCAAGAAT 4700

GTTTCCATCTTCTTTAAAATCAATACCTTTTAACTCGATTCTATTAACAAGGGTATCACCTTCAAACTTGACTTCAGCACCTGAAAATTTAAATATGTAT 4800

GGTTAGTTAGTACCGAACTGTTTAAACTTACGTGTCTTGTAGTTCCCGTCATCTTTGAAAAATATAGTTCTTTCCTGTACATAACCTTCGGGCATGGCAC 4900

TCTTGAAAAAGTCATGCCGTTTCATATGATCTGGGTATCTCGAGAAGCATTGAACACCATAACAGAAAGTAGTGACAAGTGTTGGCTGAAAATTTAAATA 5000

ATCAGGGTTAGTTAGTATATATATGTTTAAACTTACCCATGGAACAGGTAGTTTTCCAGTAGTGCAAATAAATTTAAGGGTAAGTTTTCCGTATGTTGCA 5100

TCACCTTCACCCTCTCCACTGACAGAAAATTTGTGCCCATTAACATCACCATCTAATTCAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTT 5200

TACTCATTTTTTCTACCGGTACCCTCCAAGCGTACGTTGGTGGTTGTCGGATTGAAGTCATTTTAAGTGGAACAACTGATTTTAGAGGTGGAAGCGGTGG 5300

AACGTCCGTCGTGTCGATGCTCATTGATTGGAGCTTTTCAGGTTCTGGTTTTTCAGATATCACGGGGTCTGAAATTATAAAATATGAATTTGAGATTGAG 5400

GAGCATTCTTCGAAAAATACAAATATTTTAAGATTCCTAGTTTTGACAAAAATTTATAGCCCAAGTCTGAAAATCGCCGAAAAGTCGGAAAATGCTTGGT 5500

TTCAGAAATCACACATTTAAAATTTTTAAATTTTCCTCAAAGGAATTTTTTTTCAATTTCTGGAATTTTTCTCAATTTTGTTTTTTACAATTTTTGGAAT 5600

TTTTCTCAATTTTTTTCCGGAAAATAGTGAATTTTTTGATAGTTTTTCACTATGTCTCTATTTCTGAAAATATGAATCTTTTGATATTTTTTTTGTTGAT 5700

TTTTTGAAAAGTAGCTTTATTTGTGAGTTTTCAAAATATTTTTTTTGGAATTAAATTTCGCTTTTTCTGATTTTTTGTCGTTCCTGCGAACGAAAAAATC 5800

GAATAAATTAAAATGTATTTTAAAATATGTTTTGTAAATTCACAAACAAATCTTACTATTAGACATATTTGGAAAAACATTTTTGAAAAAAATTCCCAAG 5900

AATTTAGAAATTTTTATAGAGTATTATTTTATTTTTTAACTCAATTTTTTAAACAACTTTCAGAAAAAAAACTTATGGTCAGAAAAATCAAAGTTTTACT 6000

GAACCTTCTACAAATAAAATAATTGTACTCCTCCTTACTATTCCCGCACAAACTTTAAATAATCCAAAAATTACTTCCCATCACATGGTCAACAAAAACG 6100

TCACCGACCGAAACATAAGTAAAGTTCAGAAATAAACGTACCATTTGTCAACTTCTTATGTGCCATAATTGGACTAACAACTCCGATTGTTGGCTGGGTG 6200

TCACGTGGCGGAACAGCTGGAGGTGGATCTTTTTTCGCTGTGCTTTTCACTCCTTTCACTGCCAGTGTTGGTTTTTCCTAAAAATTGATTATTTTTCGAA 6300

AATATAGGAAGGCTCTTAGGAAGGCGAAGCCAGACATTAAAGTGTATCAAAATCTTAAAATCTAAAAAAGGGAAAATCTGCAACAAAAGACGACAAAAAA 6400

GGAAGAATGGAAAGACGGAAAGCGTATTGAATTTCAAGATATTTTCCATTTTTGCACTTTTGCCTAGTCCGACTTTGAAGCCAGCTTCGAATCTTTAGAT 6500

TGCAACTACTCGGGCCAAAGAATATGTACATGGTTCAGGCCGGCCCAGTTTAGCCCATGCTTGATCAATGGCTTTGCCAGCCTCTAAACCAAGTTAATAT 6600

AATTTAGGCCAAACTTGACAAGAAGCCTCTACCCAGAAACAAAAAATACTCACCTCAACAGGCTTGTTGATAATATTTGGAGAAACCGGTGTTGCGGCTA 6700

TCGGCTGTCTGATAGCTGTCACGATGGAGGCGTTAAGAGTAAGATCGTCTGATGACGGAGACTTTTCGTCTGACGTTGAACTCTGAAAATTTTGAAGTTT 6800

TATAGAAGGGGACGGTTTCAAATGGGATAAATTGAAACTTAAAGTTTCATCTCTAAAAATTCAGTTGCACAAAAAATTAGTTTCGAACTTATAAAATGAG 6900

ATTATTAATTCGAAAGCTTCTCGGCGACCACAAAAGTTTTAGGGACGGAACTCCTATCTGAATGTTCTAAAATTTTATCCATATAAAAAAGTTTTAAGTA 7000

AAACATTTTTTCGTAAATATCTTAAACTTTTAACATCAATTTTTGATCTGAAAATAAAATTTAAAAAGTTTGGGTTTGACAAATACCGTATGACATTTAT 7100

CCTAAAAATGTATAAAGCAGGTGCTCTGCAAAGTTTTTTGTACCAACAAATGGGCGGATTCACATTTTATGGCAGGTGTCCCAGAGACCTCACTTTTATA 7200

AACATCCGAGACTATGTGAAGGTCACGAAAACAAAAAAAAACAGTTCTCCTAAACAACGAACCTTGGAAGATGTGGAATGCATGCTTAGGGAACCTTCCG 7300

TCGATGATGACGTTGGCGACGTGCTGTTGAATCCAGCGTATCCGGACTCTTCTTCACTGCTCTTTGAGCATCGTTTCGAGTCTTGTTGAGATATGATTGG 7400

GGCGTCCGTTTTACGGTAGGAAACTTTTGGCTGGAAATTACTATGTAAAAAACGAAGATGCACAGTTTTTTTTTCAAAAAAATTTTTAAAAAATGTTTTC 7500

TCATAAATAATAGAAGCCATAAGACAATTCCGGAAAGTCAAAAGAATCTCGTATATTTCAGTAGGAAGTTTTAAAAAATATAGCAGGTCTAAGAAGCTCT 7600

AATTTATATGAAAAAAATAAATACTAGAAACTGAAAACGTAAAGTTAATTTTTCAAAAAGTGCATCTCACACAAAAATGTTTTGCAATTTGTGCTCTTTC 7700

ACTCTGTAGTAGCTCATTTTGAAAAACGAATTGAAAAACCGGCGAAGAAGAGCTGAGAAATTGCTTTACTCTTTGTCAGCCAGATGTGTTTTTCGGCCCT 7800

CGACGTCGAAAAATGCTCGGAATATCACTTCCGTTGCAACATTTTTCCGCGAAGAGTCTTTATGGGGGGACACAACGACAAAAAGATGGGACATAGATTC 7900

TGAAAGAGTGTCAAGGAGAAAGAAGAAAATGGGAGACAATTCCCAATTGTCTCATCATTTTTATTAGTAGGTTGGTTTTGGTTTTGGTAGAGTCAAGGAA 8000

CGGAAGAAGTACGAATCGGGTTCTAAATATAGATGAGGCCCAGAGCTTCGCAAAACTAGGGTTTTAAAATTTCTTGAAAATCAACTTTGACAGCTTATAT 8100

CTCACTTGTCTTTGGTCCTATACAAAAATTATCAACTGTCAAAATATTTTTTATTTCTTTTTCTATATCTGATATCTTCAATATGAACCAAGATATAAAT 8200

CTTCAAAGTAAAATAGTGGGATGCAATACTATCAGAGGAAAAAATCTTATTTTTAATTTATTTTGTAGAATTTTCAGTTTTTAAAAAATACATTTCCAAT 8300

TTCCGATTTCAAATACTCACCGTACAAAGCTTCGACATAGACGTGGCACTTCCCAGCTTACTGGTCGGCGGCTTCAGGCCACTTTTCACTGGGGCAGCGA 8400

TTTTCGACAAAGTTTGTTGTTGAGGCACCGCCGCCGCCTTTCTCGTAGGTTGTGGGCTATTCGATGAGGAAGATGGGTTTTTGCTACTGAATAACTTTAA 8500

TTTCAGCATTCCACCACCACCACCACCGCTGTTATCGGGCTCTTGTTTTGCTCCAATAGTCTTCACAGAAGCAAGTTTTGGGGTGCTCACGGCTTTCGGA 8600

GCGGCTAGCTTTGAGCTTCCGATTTTTGTAGTTGTAGCAACACGAACTAGCTGGGTTTGTGGTCTAGAAGGTTTTTGGAGTTGGGAGGTAGGTCGGTTTA 8700

GATTCGAAATAGAGCTGTACGTTGATGATGATTCTGAAATAATATAATTTCTAAAAAGAAGCCGGAAGGATCGGATACCTAAGCTCTTCGCAGATGTGGA 8800

TATCGTCGAGCCAACATTATTATTGCCACTCGAACGGCTCGACGGACGGAATGAATTTGTATTATTTGATGAAGTGGTTGATGATGAGGGTGGTTTAAGT 8900

CCAGACGTCTTTGGCTTGATACCAATCTTTGATGAATCTGTAATTGAAAGTGTGAGAGAAAGGGATAACGATTTTACGCATCACAGGGCATTAAGATTTT 9000

donderdag, 27 november 1997 16:52
fig 27 pNP3 Map (1 > 13621) Site and Sequence Page 4

AAGTAACAAGAAACTGGAAACATTTAAAATTGAAAACTCTCTTGGAAATCTAGTCTGGATTAGACTGAAGCAAGCAAGTCAATCTAGGCTTAAACTAAAT 9100
GCAAGTCATTTCTGCTTGGGCTTGAGCTTATGCGCAACTAAGATCTAGAATAAGGCTATAGTTCAGGCTTAGGATTTCCGAGTTCAATCTAAACGTTACA 9200
CTAGATTTATGCAGGCACTTCAGTTTTCTGTTCAGCCGAATCTTGCCTAAACATCTTGAAAACCTTTAAATAGAATTTTGATCTGGCGGTTCCTATTTGT 9300
TTTGTTTTTTTCTTCTTTTCCAAATCTACATCTACAACCTCATTAATTTTAATATTTTTAAATTGCCTCGTTTTTTATGTATTGCATAAAGATAAGCTCA 9400
TAAATATACTAAATTTTCCCGGCTTTCAGGCAGCCACAAAAAATAGGCGCGAGTCAAAACGAAGAAGAAAACAAGGAGGTCTATGGCTATTTGAAGAGTT 9500
AGTATGAGCGTGAATTTGAGACCGAACGGACGAGATGAAAAGGGTCACACGCCATGTGGCTCAGTATCTCCTTCTCTTAAAAAGTAGAGTCGGGCTTTTG 9600
CTATACTTTTTATGATACGAGTGAGATGGATGAAGAGGTGAGAAAGTTTTTTAATGGGCCAAAAAGCATGGAAAATTGTACGAACTGGCTGAATGTTCAA 9700
AGATTGAATTTGATAGATAATCTGTGCTTGAATTAAAATTCTAACTGCTTGCACTTTATTACGCGTGCCCAATAATTATCAACCGTCTAGCTTAATTTTT 9800
GGATATGTGCTGAGTAGCGATTCATAGCATGGGAAATGTGACACGGAGCTTCCTATAACTTGGGGGATTCAAGGTTTTGAGATGGAAAACAGTGATACAT 9900
ATAAAGATATTGTCAACAGTAGAAATAAAATTTTAGCCATACAATTTGTTAGGGAATATTTATTGATTTTTTGAAACTTTTCAGATTTTTTAAAATCGAA 10000
AATATCGTTTTTTTAAATAAACTTTTAAAAGTCGTGTTTATCCCTAATTGCACAAAATTAAATATCAAAATAAATGGCTTGAATAGTAGTAAAATGTATT 10100
CCAGTTTGTGGTAACCCAGAATGAGTGGCGGAATAGGAAAGAGCGCATAAAACCCGACATTAATTGTCAGTGATGAGAGCGGGGGAAGACTCATGCGCAA 10200
GTTGAGACGATGAAAAAGCAATTCTACACTCATACTACTACCAGCCAGTGTCATCGGCTCTATCTTTATTCATCAATTTCCGAAAGAGTAATTGCCATGC 10300
TCTCATTCGAATTAATTGACCGTCTCCCCGTGTAGTTTCCTTCCACACACAAATGTTTATAAGCAAAATGGACAACGAAATTGGAAAATTGTTGTTGTTG 10400
GAGATGGAACGGGAGGAGGAGGCGGGAGAAGGAGGGAGAAGAGGGACAACCTGTCCCTGTCAATTTAGTTAGCCTCTTCTGTGTTTTCTCACATCAATTT 10500
AAAAGTCATTATGAAGCCGAGAAATTGATTTTGGTGGGCACTTTTCGGGCAAGGGGAGGCGATTTGTAAATTGCAAAATTTTATAAAAATTAAGAGTTGA 10600
AACATGGTGGCGGTGAGGGTTTGAAATAGTTTTTTTTTACTTTTTCATACAAAAAGGAAGGTTCTGAAACTAGTGCGAATTTGAAAAAATCTTCAAAACA 10700
TGTCTTAAACATTTGAGAATAGCACAAATTTGTTTCAGATCATTATTTCCAGTATAGACTTGCGCTCCTTCCACTTTAACAGCTTGTATCTAGGTTCCAT 10800
TTGGATCTATTACAAAGCTGTCAACTGACAAAATTTGTGCAAAATACTTTTCTACATTTGTATAGTGGAAAATTTTTTGATAAAACTCAAGAAAATTGAG 10900
ATATAAGCCGTCATATTAGAGCAATAAAGGTGGAAGTTTTGCAAAAAAACTACATTTTTATCATCTTCTTTTCTGGGCATTTTATACAAGTTTGAGCTCA 11000
AAATAAAATTCTTACCAATTTTCGATATTCTTGACTGTGGAGTCTGAAGCCTGGATGTTGACATTTGTGGAAAGTTGGAATTTGGGTTAGTTGCTGAAGC 11100
GGTTGCTGACGTGGCGACACGTGGCGAGGGTAATCTGAAAATGGAATTGTATTGCAAGTTTTGTAAGATCTAGTTGATCGAAAAACTAAATCAAAGTTAG 11200
GGCAAATAAGCAAAGTAAATAATGTTCCTTTTAATATTTTCTCAATGAAAAATCACTCAACAACTAGTCATACAGAATAGACTCAAGTCGAAGATAGTTA 11300
TTAAAGAACACATTTTTGTAGTCGTAAACTCAAAATTAAACTCACTTAGAAACCGCGGGTGGCATAATGGATGTGGGTAGTTGCTCCAATTTCTTCTGAT 11400
CTCGAGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTT 11500
ATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC 11600
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCT 11700
TCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA 11800
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA 11900
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT 12000

GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT 12100

AGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG 12200

ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTAC 12300

GGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG 12400

CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA 12500

GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC 12600

TAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCT 12700

GACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTT 12800

ATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGA 12900

GTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCG 13000

GTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC 13100

AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA 13200

TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG 13300

GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTT 13400

TACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTC 13500

CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA 13600

CATTTCCCCGAAAAGTGCCAC 13621

FIGURE 28.

SIGNATURE SEQUENCES :

Different signatures can be used to define to identify the UNC-53 gene family :

Aminoacids are listed in one letter code
X equals any aminoacid
X(3,5) equals 3 to 5 X's
(D,E) means D or E at a given position

The signatures should be used to screeen a database using a weight matrix of conservative substitutions.

BLOCK A :
GSXLSLASX(3,5)YXSXXEX(4,5)IRXXXR(D,E)LEXXXXVXXLTXXXXXXXXLXXXFEQXL

BLOCK B :
KXKKSWXXXXXXXXFXK

BLOCK C :
LARGE FAMILY :
VXXL(K,R)XELX(D,E)(R,K)(D,E)XXLXX(V,I)RL(D,E)XLXXAXXXDXLRE(T,A)XXXXXXEXXXLKXEXD(R,K)LX

VERTEBRATE FAMILY :
VXXLRXELX(D,E)(R,K)(D,E)MKLTDIRLEALXSAHQLDQLLREXMXXMQXEXXXLKAENDRLK

BLOCK D :
LARGE FAMILY :
W(K,D)X(I,L)DXX(I,V)XX(L,V)F(K,E)XY(I,V,L)XXXDXXXXLG(I,L)X(2,3)(D,E)S(I,V)XGYXI(G,S)(E,H)(L,V,I)(R,K)(R,K)

VERTEBRATE FAMILY :
FXXGCXXVSGKXXWXXLDXXVXX(L,V)FK(D,E)YIXXXDPXXXLG(I,L)XX(D,E)S(I,V)XGYSI(G,S)XXKR

BLOCK E :
GXXGXGKS/T
and
F(K,R)MXXXSNX(3,8)GF(I,L,V)(I,L,V)(R/K)Y(I,L,V)(R,K)(R,K)(R,K)XV(D,E)
and
F(I,L)EKXXXX(D,E)XXXGPXXX(L,I)XCPXXXXXX(R,K)XWFIXLWNXXXIPY(L,I)XXXA(R,K)DGX(K,R)XXGXXXX(F,W)EDP

Block F

(W/F) (D/E) DSSS (V/L/I) SSGISD (T/N)

Tuesday, 18 November 1997 10:34 Page 1
fig 29 pEGFPsac (1 > 5100) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

7p.

Bgl I

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG 100
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC
. L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

Aat II Aat II

CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT 200
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

Bgl I Nde I Aat II Bgl I

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA 300
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

SnaB I Nco I

CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA 400
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W

Aat II

TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA 500
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .

Nhe I Ec47

ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA 600
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L

Nco I

CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG 700
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC 800
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T

Tuesday, 18 November 1997 10:34

fig 29 pEGFPsac (1 > 5100) Site and Sequence


```
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC  900
 T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E

                                       KspI
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC  1000
 R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT  1100
 D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC  1200
 V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT  1300
 P  D  N  H  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G

                            BspM II   Bgl II                                                    Asu II
                                                                                                EcoN I
TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG  1400
                                                      C.e.unc53 sac
 I  T  L  G  M  D  E  L  Y  K  S  G  L  R  S  T  S  N  V  E  L  I  P  I  Y  T  D  V  A  N  R  H  L  S

                                         Nru I                                               EcoR I
GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT  1500
                                          C.e.unc53 sac
 K  G  S  L  S  K  S  I  R  D  I  S  N  D  F  R  D  Y  R  L  V  S  Q  L  I  N  V  I  V  P  I  N  E
```

Tuesday, 18 November 1997 10:34 Page 3
fig 29 pEGFPsac (1 > 5100) Site and Sequence

```
     Bsm I
TTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT  1600
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTGTAGCTTGGACCTACCGGAGCTTTGCACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA
                                   C.e.unc53 sac
F  S  P  A  F  T  K  R  L  A  K  I  T  S  N  L  D  G  L  E  T  C  L  D  Y  L  K  N  L  G  L  D  C

                                                               Ear I
                EcoR V                                Pvu II   Ksp632I                Hind III
CGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA  1700
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCGTCTTCGAAGCCGTTGACTT
                                   C.e.unc53 sac
S  K  L  T  K  T  D  I  D  S  G  N  L  G  A  V  L  Q  L  L  F  L  L  S  T  Y  K  Q  K  L  R  Q  L  I

                                                       Sst II
                                                          Xma I
                                                           Apa I
                                                            Sma I
                                                              Bam HI        Xba I       Bcl I
AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACC  1800
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCCGGGCCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGG
                 C.e.unc53 sac
K  D  Q  K  K  L  E  Q  L  P  T  S  I  M  P  P  A  G  P  G  S  T  G  S  R  .  L  I  I  I  S  H  T

                    Dra I                                                   Bsm I        Hpa I
ACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTG  1900
TGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAAC
T  F  V  E  V  L  L  A  L  K  N  L  P  H  L  P  L  N  L  K  H  K  M  N  A  I  V  V  V  N  L  F  I

                                                                       Bsm I
CAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAA  2000
GTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTT
A  A  Y  N  G  Y  K  .  S  N  S  I  T  N  F  T  N  K  A  F  F  S  L  H  S  S  C  G  L  S  K  L  I  N

     Mlu I               Ssp I
TGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGG  2100
ACATAGAATTGCGCATTTAACATTCGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCC
V  S  .  R  V  N  C  K  R  .  Y  F  V  K  I  R  V  K  F  L  L  N  Q  L  I  F  .  P  I  G  R  N  F
```

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence


```
                                                   BsrI
CAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTC
GTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAG  2200
Q  N  P  L  .  I  K  R  I  D  R  D  R  V  E  C  C  S  S  L  E  Q  E  S  T  I  K  E  R  G  L  Q  R

                                        Dra III
AAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGA
TTTCCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCT  2300
Q  R  A  K  N  R  L  S  G  R  V  P  T  T  .  T  I  T  L  I  K  F  F  G  V  E  V  P  .  S  T  K  S  E

                                         Nae I
ACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCT
TGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGA  2400
P  .  R  E  P  P  I  .  S  L  T  G  K  A  G  E  R  G  E  K  G  R  E  E  S  E  R  S  G  R  .  G  A

                                           KspI                                                  KspI
GGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGC
CCGTTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCG  2500
G  K  C  S  G  H  A  A  R  N  H  H  T  R  R  A  .  C  A  A  T  G  R  V  R  V  H  F  S  G  K  C  A

                                                BspH I                              Ssp I    Ear I
                                                                                             Ksp632I
GGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
CCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCT  2600
R  N  P  Y  L  F  I  F  L  N  T  F  K  Y  V  S  A  H  E  T  I  T  L  I  N  A  S  I  I  L  K  K  E  E

  OxaN I           Pvu II                                                                    Sph I
                                                                                               Ava III
                                                                                               Nsi I
GTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTC
CAGGACTCCGCCTTTCTTGGTCGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAG  2700
S  .  G  G  K  N  Q  L  V  N  V  C  Q  L  G  C  G  K  S  P  G  S  P  A  G  R  S  M  Q  S  M  H  L

                                                                Sph I
                                                                  Ava III
                                                                  Nsi I
AATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
TTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGG  2800
N  .  S  A  T  R  C  G  K  S  P  G  S  P  A  G  R  S  M  Q  S  M  H  L  N  .  S  A  T  I  V  P  P

                              BsrI                 Nco I                                         Bgl I
                                                                                                 Sfi I
TAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTC
ATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAG  2900
L  T  P  P  I  P  P  L  T  P  P  S  S  A  H  S  P  P  H  G  .  L  I  F  F  I  Y  A  E  A  E  A  A  S
```

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 5

```
                                                         Stu I
                                                          Avr II                 Cla I
GGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT  3000
CCGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTA
 A  S  E  L  F  Q  K  .  .  G  G  F  F  G  G  L  G  F  C  K  D  R  S  R  D  R  M  R  I  V  S  H  D

          BspM I           Xma III
TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCC  3100
ACTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGG
 .  T  R  V  I  A  R  R  F  S  G  R  L  G  G  E  A  I  R  L  .  L  G  T  T  D  N  R  L  L  .  C  R

                            Nar I
                              Bbe I                                                       KspI
GTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGT  3200
CACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCA
R  V  P  A  V  S  A  G  A  P  G  S  F  C  Q  D  R  P  V  R  C  P  E  .  T  A  R  R  G  S  A  A  I  V

                            Fsp I
     Bal I                     Pvu II        Tth I
GGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT  3300
CCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGA
 A  G  H  D  G  R  S  L  R  S  C  A  R  R  C  H  .  S  G  K  G  L  A  A  I  G  R  S  A  G  A  G  S

                                                                                             BspM I
CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCAC  3400
GGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTG
P  V  I  S  P  C  S  C  R  E  S  I  H  H  G  .  C  N  A  A  A  A  Y  A  .  S  G  Y  L  P  I  R  P

                                                                         Ear I
                                                                         Ksp632I
CAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCG  3500
GTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGC
P  S  E  T  S  H  R  A  S  T  Y  S  D  G  S  R  S  C  R  S  G  .  S  G  R  R  A  S  G  A  R  A  S  R

                                Sph I                      Nco I
AACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCG  3600
TTGACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGC
 T  V  R  Q  A  Q  G  E  H  A  R  R  R  G  S  R  R  D  P  W  R  C  L  L  A  E  Y  H  G  G  K  W  P

                             Nae I          Rsr II                                            Ear I
                                                                                              Ksp632I
CTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAA  3700
GAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTT
L  F  W  I  H  R  L  V  P  A  G  C  G  G  P  L  S  G  H  S  V  G  Y  P  .  Y  C  .  R  A  V  R  R
```

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 6

```
TGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
ACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGA 3800
M G . P L P R A L R Y R R S R F A A H R L L S P S . R V L L S G T L

 Asu II                                      BspM I
GGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTC
CCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAG 3900
G F E M T D Q A T P N L P S R D F D S T A A F Y E R L G F G I V F

  Nae I                    KspI                         Avr II
CGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
GCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATG 4000
R D A G W M I L Q R G D L M L E F F A H P R G R L T E T R K E T I

          KspI
CGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCA
GCCTTCCTTGGGCGCGATACTGCCGTTATTTTTCTGTCTTATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGT 4100
P E G T R A M T A I K R Q N K T H G V G S F V H K R G V R S Q G W H

CTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTC
GAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAG 4200
S V D T P P R P H W G Q Y A R V S S F S P P H P P S S G E G P G L

                        Alwn I     OxaN I                      Dra I              Dra I
GCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGG
CGTCGGTTGCAGCCCCGCCGTCCGGGACGGTATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCC 4300
A A N V G A A G P A I A S G Y S Y I L . I D L K L H F . F K R I .

                BspH I
TGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
ACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAAC 4400
V K I L F D N L M T K I P . R E F S F H . A S D P V E K I K G S S .

AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTT
TCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAA 4500
D P F F L R V I C C L Q T K K P P L P A V V C L P D Q E L P T L F

          BsrI
CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTA
GGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGAT 4600
P K V T G F S R A Q I P N T V L L V . P . L G H H F K N S V A P P
```

Tuesday, 18 November 1997 10:34 Page 7
fig 29 pEGFPsac (1 > 5100) Site and Sequence

```
                                                Alwn I
CATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
                                                                                                      4700
GTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCG
 T  Y  L  A  L  L  I  L  L  P  V  A  A  A  S  G  D  K  S  C  L  T  G  L  D  S  R  R  .  L  P  D  K  A

                                          ApaL I
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGC
                                                                                                      4800
CGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCG
  Q  R  S  G  .  T  G  G  S  C  T  Q  P  S  L  E  R  T  T  Y  T  E  L  R  Y  L  Q  R  E  L  .  E  S

GCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGT
                                                                                                      4900
CGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCA
  A  T  L  P  E  G  R  K  A  D  R  Y  P  V  S  G  R  V  G  T  G  E  R  T  R  E  L  P  G  G  N  A  W

ATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
                                                                                                      5000
TAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCG
 Y  L  Y  S  P  V  G  F  R  H  L  .  L  E  R  R  F  L  .  C  S  S  G  G  R  S  L  W  K  N  A  S  N  A

                                                                                                 Ava III
                                                                                                 Nsi I
GGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
                                                                                                      5100
CCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
  A  F  L  R  F  L  A  F  C  W  P  F  A  H  M  F  F  P  A  L  S  P  D  S  V  D  N  R  I  T  A  M  H
```

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

PAGE 1 ~~Page 1~~ 16p

```
                                                                                    Bgl I
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC  100
 .  L  L  I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  V  P  A  V  L  T

                            Aat II                                                  Aat II
CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATGCCA  200
A  Q  R  P  P  P  I  D  V  N  N  D  V  C  S  H  S  N  A  N  R  D  F  P  L  T  S  M  G  G  V  F  T  V

              Bgl I                Nde I                       Aat II                   Bgl I
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT  300
 N  C  P  L  G  S  T  S  S  V  S  Y  A  K  Y  A  P  Y  .  R  Q  .  R  .  M  A  R  L  A  L  C  P  V

                                          SnaB I             Nco I
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT  400
H  D  L  M  G  L  S  Y  L  A  V  H  L  R  I  S  H  R  Y  Y  H  G  D  A  V  L  A  V  H  Q  V  A  W

                                                Aat II
TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT  500
I  A  V  .  L  T  G  I  S  K  S  P  P  H  .  R  Q  V  E  F  V  L  A  P  K  S  T  G  L  S  K  M  S  .

                                                                                       Nhe I Ec47
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT  600
 Q  L  R  P  I  D  A  N  G  R  .  A  C  T  V  G  G  L  Y  K  Q  S  V  F  S  E  P  S  D  P  L  A  L

         Nco I
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAAGTCGC  700
                                        eGFP.C.e.unc53
P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S

TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG  800
                                 eGFP.C.e.unc53
V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T
```

Tuesday, 18 November 1997 10:34 Page 2

fig 30 pEGFP72 (1 > 9697) Site and Sequence

CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG 900
GTGGGACTGGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTC

eGFP.C.e.unc53

T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

KspI

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG 1000
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC

eGFP.C.e.unc53

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA 1100
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTGCAGATATAGTACCGGCTGTTCGTCTTCTTGCCGTAGTT

eGFP.C.e.unc53

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG 1200
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGTCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC

eGFP.C.e.unc53

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA 1300
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT

eGFP.C.e.unc53

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

BspM II Bgl II Asu II EcoN I

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC 1400
AGTGAGAGCCGTACCTGCTCGACATGTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG

eGFP.C.e.unc53

C.e. unc53

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

Nru I EcoR I

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA 1500
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT

eGFP.C.e.unc53

C.e. unc53

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Bsm I

TTCTCGCCTGCATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTGTAGCTTGGACCTACCGGAGCTTTGCACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA 1600

eGFP.C.e.unc53

C.e. unc53

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

EcoR V, Pvu II, Ear I, Ksp632I, Hind III

CGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCGTCTTCGAAGCCGTTGACTT 1700

eGFP.C.e.unc53

C.e. unc53

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L K

Sst II, PmaCI, PmaCI

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTCAGCAACCGCTTCA
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT 1800

eGFP.C.e.unc53

C.e. unc53

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT 1900

eGFP.C.e.unc53

C.e. unc53

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

Aat II

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCCGTCGAGCCGTTCGAGTGGCAATAATAA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT 2000

eGFP.C.e.unc53

C.e. unc53

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

Tuesday, 18 November 1997 10:34

fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4

EcoR V | Ear I / Ksp632I | Asu II

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA 2100

ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTTGGCTGGATGGAGGGTTGAGGTTTTT

eGFP.C.e.unc53

C.e. unc53

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q K

Xba I | Nhe I

CCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAAACTTG 2200

GGAAGATCTGGTGTTTGGGTCGATCAAGCACAACGATGTTGATGTTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGGTTTTGAAC

eGFP.C.e.unc53

C.e. unc53

P S R P Q T Q L V R V A T T T K I G S S K L A A P K A V S T P K L

Bsm I

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGGAATGCTGAAATTAAAGTTATTCAGTAGCAAAAACCCATC 2300

GAAGACACTTCTGATAACCTCGTTTTGTTCTCGGGCTATTGTCGCCACCACCACCACCACCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG

eGFP.C.e unc53

C.e. unc53

A S V K T I G A K Q E P D N S G G G G G G M L K L K L F S S K N P S

TTCCTCATCGAATAGCCCACAACCTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG 2400

AAGGAGTAGCTTATCGGGTGTTGGATGCTCTTTCCGCCGCCGCCACGGAGTTGTTGTTTGAAACAGCTTTTAGCGACGGGGTCACTTTTCACCGGACTTC

eGFP.C.e.unc53

C.e. unc53

S S S N S P Q P T R K A A A V P Q Q Q T L S K I A A P V K S G L K

BstX I | Hind III

CCGCCGACCAGTAAGCTGGGAAGTGCCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTTCCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAG 2500

GGCGGCTGGTCATTCGACCCTTCACGGTGCAGATACAGCTTCGAAACATGCGGTTTTCAAAGGATGGCATTTTGCCTGCGGGGTTAGTATAGAGTTGTTC

eGFP.C.e.unc53

C.e. unc53

P P T S K L G S A T S M S K L C T P K V S Y R K T D A P I I S Q Q

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Ear I
Ksp632I BspM II

```
ACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTTCTCAGGCCTATGCGACCTAAGTTGTCGTGCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTCGTA  2600
```

eGFP.C.e.unc53

C.e. unc53

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

Bsm I
Sph I
Ava III
Nsi I

```
GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCCG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTCGATAGTCTGTCGGC  2700
```

eGFP.C.e.unc53

C.e. unc53

H S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

Ssp I

```
ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTCGGACAACTCCTTTTTGGTTGTGACCGTCACTTTCCTCACTTTTCGTGTCGCTTTTTTCTAG  2800
```

eGFP.C.e.unc53

C.e. unc53

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

PmaCI
Pvu II PmaCI EcoR V

```
CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTCGACAAGGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGTATTCTTCAACTGTTTACTGGGGCACTATAGACT  2900
```

eGFP.C.e.unc53

C.e. unc53

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Alwn I

```
AAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA
TTTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAGT  3000
```

eGFP.C.e.unc53

C.e. unc53

K P E P E K L Q S M S I D T T D V P P L P P L K S V V P L K M T S

Tuesday, 18 November 1997 10:34 Page

fig 30 pEGFP72 (1 > 9697) Site and Sequence

```
               SplI
                SplI
ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACT 3100
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTTAGTGTAGCGGACAGTTCAGCAAACCTATACTCGTCAGCAGGCGCAGACTTCTGA
                                      eGFP.C.e.unc53
                                      C.e. unc53
 I  R  Q  P  P  T  Y  D  V  L  L  K  Q  G  K  I  T  S  P  V  K  S  F  G  Y  E  Q  S  S  A  S  E  D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC 3200
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTTTGAAGACCATTAGTAAGCGACCTCTCTTCCTACCCTTTCTTATTCTGTAGTCTTAG
                                      eGFP.C.e.unc53
                                      C.e. unc53
S  I  V  A  H  A  S  A  Q  V  T  P  P  T  K  T  S  G  N  H  S  L  E  R  R  M  G  K  N  K  T  S  E  S

CAGCGGCTACACCTCTGACGCCGGTGTTGCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT 3300
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTTACTCCCTCTTCGACTTTCTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA
                                      eGFP.C.e.unc53
                                      C.e. unc53
 S  G  Y  T  S  D  A  G  V  A  M  C  A  K  M  R  E  K  L  K  E  Y  D  D  M  T  R  R  A  Q  N  G  Y

        Asu II                                              Sst I                      BspM II
CCTGACAACTTCGAAGACAGTTCCTCCTTGTCGTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGG 3400
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCCTGCTAAACAGGCCTCATCTGTACC
                                      eGFP.C.e.unc53
                                      C.e. unc53
P  D  N  F  E  D  S  S  S  L  S  S  G  I  S  D  N  N  E  L  D  D  I  S  T  D  D  L  S  G  V  D  M

CAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTCGCCATCCCACGTCTTCTTCCTCAAAGCCCCGAGTCCCCAGTCGGTCCTCCACATCAGT 3500
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAAACAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA
                                      eGFP.C.e.unc53
                                      C.e. unc53
A  T  V  A  S  K  H  S  D  Y  S  H  F  V  R  H  P  T  S  S  S  S  K  P  R  V  P  S  R  S  S  T  S  V
```

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 7

Xho I, Bgl I, Nar I, Bbe I

```
CGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGTCCCAGTGCCGAACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGGACAA
GCTAAGAGCTAGAGCTCGTCTTGTCCTCTTACACATGTTTGAAGACAGGGTCACGGCTTGCTCGGTTGCACCGCGGCGACGGTGGAGTTGGAAGCCTGTT  3600
```

eGFP.C.e.unc53

C.e. unc53

D S R S R A E Q E N V Y K L L S Q C R T S Q R G A A A T S T F G Q

Xma I, Sma I, Pvu II, Spe I, Sal I

```
CATTCGCTAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTGTCAGCTGATAAGGACACAATGTCTATGCACTCACAGACTAGTCGACGAC
GTAAGCGATTCTAGGGGCCCTATGAGTAGGATAAGAGGTGTGAATAGTCACAGTCGACTATTCCTGTGTTACAGATACGTGAGTGTCTGATCAGCTGCTG  3700
```

eGFP.C.e.unc53

C.e. unc53

H S L R S P G Y S S Y S P H L S V S A D K D T M S M H S Q T S R R

```
CTTCTTCACAAAAACCAAGCTATTCAGGCCAATTTCATTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATCCACCGAGCACAGAATGGCGGCTCT
GAAGAAGTGTTTTTGGTTCGATAAGTCCGGTTAAAGTAAGTGAACTAGCATTTACGGTGGAAGTTCTCAAGTGTAGGTGGCTCGTGTCTTACCGCCGAGA  3800
```

eGFP.C.e.unc53

C.e. unc53

P S S Q K P S Y S G Q F H S L D R K C H L Q E F T S T E H R M A A L

Bam HI

```
CTTGAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCCTACTCGGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGA
GAACTCGGGCTCTGCCCACGGCTTGAGCTACAGCTTTATACTAAGAAGTCCTAGGATGAGCCGCGCAAGGGCTCCACCTTCGAGATGACCATAGATACCT  3900
```

eGFP.C.e.unc53

C.e. unc53

L S P R R V P N S M S K Y D S S G S Y S A R S R G G S S T G I Y G

Bam HI, Nhe I, Nde I

```
GAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCCAAAAGTGAGATGGGATCCCAACTATCACTGGCTAGCACGACAGCAT
CTCTGCAAGGTTGACGTGTCTGATAGGCTACTTTTTAGGGGGCGTGTAAGACGGTTTTCACTCTACCCTAGGGTTGATAGTGACCGATCGTGCTGTCGTA  4000
```

eGFP.C.e.unc53

C.e. unc53

E T F Q L H R L S D E K S P A H S A K S E M G S Q L S L A S T T A

Tuesday, 18 November 1997 10:34 Page 8

fig 30 pEGFP72 (1 > 9697) Site and Sequence

PmaCI PmaCI Sal I

ATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACACTGTCGACTCACTAACCAAGAAACAGGA 4100
TACCTAGAGAGTTACTCTTCATGCTTGTACGATAAGCCCTGTACCGTGCACTGAACCTCACAATGTTCTTGTGACAGCTGAGTGATTGGTTCTTTGTCCT

eGFP.C.e.unc53

C.e. unc53

Y G S L N E K Y E H A I R D M A R D L E C Y K N T V D S L T K K Q E

Hind III Cla I Ear I Ksp632I

GAACTATGGAGCATTGTTTGATCTTTTTGAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACGATTC 4200
CTTGATACCTCGTAACAAACTAGAAAAACTCGTTTTCGAATCTTTTGAGTGAGTTGTGTAACTAGCTAGGTTGAACTTCGGACTTCTCCGTTATGCTAAG

eGFP.C.e.unc53

C.e. unc53

N Y G A L F D L F E Q K L R K L T Q H I D R S N L K P E E A I R F

AGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCTTCGTCAACCATCTC 4300
TCCGTCCTGTAACGAGTAAACTCCCTATAATCGTTAGTAGAACGTAGGTTGAGTCGAGTACGATTGCTTCCGCGACCACTCGAAGAAGCAGTTGGTAGAG

eGFP.C.e.unc53

C.e. unc53

R Q D I A H L R D I S N H L A S N S A H A N E G A G E L L R Q P S

Cla I Cla I Sst I Ear I Ksp632I

TGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGCTCGTTTGGCAAGAACAAGAAGAG 4400
ACCTTAGTCAACGTAGGGTAGCTAGTAGCTACAGTAGCAGCAGCTTTTCGTCGTTCGTCCTCTTCTAGTCGAACTCGAGCAAACCGTTCTTGTTCTTCTC

eGFP.C.e.unc53

C.e. unc53

L E S V A S H R S S M S S S S K S S K Q E K I S L S S F G K N K K S

Bam HI Nde I BspM II

CTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAACTCTTGAC 4500
GACCTAGGCGAGGAGTGAGAGGTTCAAGTGGTTCTTCTTCTTGTTCTTGATGCTGCTTCGTGTATACGGTAGTTAAAGGCCTAGAGTTCCTTGAGAACTG

eGFP.C.e.unc53

C.e. unc53

V I R S S L S K F T K K K N K N Y D E A H M P S I S G S Q G T L D

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 9

Sst I ApaL I

AACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTTACGAAGTCCGCCTTGACAATCTGGATCGTGCCCGCGAAGTTGATG 4600
TTGTAACTACACTAACTCAACTTCGTTCTCGAGTTTCTTGCGCTATCACGTGAAATGCTTCAGGCGGAACTGTTAGACCTAGCACGGGCGCTTCAACTAC

eGFP.C.e.unc53

C.e. unc53

N I D V I E L K Q E L K E R D S A L Y E V R L D N L D R A R E V D

TTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCG 4700
AAGACTCCCTCTGTCACTTGTTCAACTTTTGGCTCTTGTTCGTTAATTTCTTTCTTCACCTGTTTGAGTGGTTGCCAGGTCGGTGAGCACGAAGAAGGGC

eGFP.C.e.unc53

C.e. unc53

V L R E T V N K L K T E N K Q L K K E V D K L T N G P A T R A S S R

KspI BsrI Asu II

CGCCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAAC 4800
GCGGAGTTAAGGTCAATAGATGCTGCTACTCGTACAGATACTACGTCGCACATCGTCATGTAGTCGATCAGTTAGAAGCTTTGCTAGGAGACCGACGTTG

eGFP.C.e.unc53

C.e. unc53

A S I P V I Y D D E H V Y D A A C S S T S A S Q S S K R S S G C N

Pvu I Hpa I EcoR V

TCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCA 4900
AGTTAGTTCCAATGACATTTGCACCTGTAGCGACCTCTTTAGTCAAGCTAGCAATTGGGCCTGTTTCTCTATTAGCATCCTATAGAACGGTACAGTTGGT

eGFP.C.e.unc53

C.e. unc53

S I K V T V N V D I A G E I S S I V N P D K E I I V G Y L A M S T

Cla I

GTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATCAACTTGGAATCGATGCTCG 5000
CAGTCAGTACGACCTTTCTGTAACTACAAAGATAAGATCCTGATAAACTTCAGATGGATAGGTCTTAACTACACCTCGTAGTTGAACCTTAGCTACGAGC

eGFP.C.e.unc53

C.e. unc53

S Q S C W K D I D V S I L G L F E V Y L S R I D V E H Q L G I D A R

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 10

Mlu I

TGATTCTATCCTTGGCTATCAAATTGGTGAACTTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCATCCAACTGACATTCTTACTTCCTCA 5100
ACTAAGATAGGAACCGATAGTTTAACCACTTGAAGCTGCGCAGTAACCTCTGAGGTGTTGGTACTATTGGTCGGTAGGTTGACTGTAAGAATGAAGGAGT

eGFP.C.e.unc53

C.e. unc53

D S I L G Y Q I G E L R R V I G D S T T M I T S H P T D I L T S S

ACTACAATCCGAATGTTCATGCACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACTCGTCA 5200
TGATGTTAGGCTTACAAGTACGTGCCACGGCGTGTCTCAGCGCATCTGTCAGACCAGGAACTATACGAAGAAGGTTTCGTTTACTAAGAGGTTGAGCAGT

eGFP.C.e.unc53

C.e. unc53

T T I R M F M H G A A Q S R V D S L V L D M L L P K Q M I L Q L V

Aat II Bsrl Bsrl Asu II

AGTCAATTTTGACAGAGAGACGTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTATGTATCTATTCGAAC 5300
TCAGTTAAAACTGTCTCTCTGCAGACCACAATCGACCTCGTTGACCTTAACCTTTCTCGTTTGACCGCTTCTGGGACCGACGAATACATAGATAAGCTTG

eGFP.C.e.unc53

C.e. unc53

K S I L T E R R L V L A G A T G I G K S K L A K T L A A Y V S I R T

Bsm I Xmn I Bgl II

AAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAACGACGCCTGGAAAAGATCTTGAGAAGC 5400
TTTAGTTAGGCTTCTATCATAACAATTATAGTCGTAAGGACTTTTGTTATTTCTTCTTAACGAAGTTCACCTTGCTGCGGACCTTTTCTAGAACTCTTCG

eGFP.C.e.unc53

C.e. unc53

N Q S E D S I V N I S I P E N N K E E L L Q V E R R L E K I L R S

Ava III
Nsi I Xba I

AAAGAATCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTTGTTGTATCCGTTTTTGCAAATGTCCCACTTCAAAACAACGAAGGTC 5500
TTTCTTAGTACGTAGCATTAAGATCTATTATAGGGTTTCTTAGCTTAACGTAAACAACATAGGCAAAAACGTTTACAGGGTGAAGTTTTGTTGCTTCCAG

eGFP.C.e.unc53

C.e. unc53

K E S C I V I L D N I P K N R I A F V V S V F A N V P L Q N N E G

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 11

EcoR V

CATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAAATGTCAGTAATGTCGAATCGTCTCGAAGGATTCA
GTAAACATCATACGTGTCAGTTGGCTATAGTTTAGGGACTCGAAGTTTAAGTGGTGTTAAAGTTTTACAGTCATTACAGCTTAGCAGAGCTTCCTAAGTA 5600

eGFP.C.e.unc53

C.e. unc53

P F V V C T V N R Y Q I P E L Q I H H N F K M S V M S N R L E G F I

Sst I, Ear I, Ksp632I

CCTACGTTACCTCCGACGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCT
GGATGCAATGGAGGCTGCTGCCCGCCATCTCCTACTCATAGCAGATTGACATGTCTACGGTAGTCTCGAGAAGTTTTAGTAACTGAAGAAGGGTTATCGA 5700

eGFP.C.e.unc53

C.e. unc53

L R Y L R R R A V E D E Y R L T V Q M P S E L F K I I D F F P I A

Ear I, Ksp632I, EcoR I, Sph I, Bam HI

CTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACTGTCCTCTAACTGTCGATGGATCCC
GAAGTCCGGCAGTTATTAAAATAACTCTTTTGCTTAAGACAACTACACTGTCAACCAGGTTCTCGTACGAACTTGACAGGAGATTGACAGCTACCTAGGG 5800

eGFP.C.e.unc53

C.e. unc53

L Q A V N N F I E K T N S V D V T V G P R A C L N C P L T V D G S

GTGAATGGTTCATTCGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCGCTGCACTTCCTTCGA
CACTTACCAAGTAAGCTAACACCTTACTCTTGAAGTAAGGTATAAACCTTGCACAACGATCTCTACCGTTTTTTTGGAAGCCAGCGACGTGAAGGAAGCT 5900

eGFP.C.e.unc53

C.e. unc53

R E W F I R L W N E N F I P Y L E R V A R D G K K T F G R C T S F E

Bam HI, Tth I, Tth I

GGATCCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAAGACCTCGTCCCGTCA
CCTAGGGTGGCTGTAGCAGAGATTTTTTACCGGCACCAAGCTACCACTTTTGGGCCTCTTACACGAGTTTGCAGAAGTTGAGGTTCTGGAGCAGGGCAGT 6000

eGFP.C.e.unc53

C.e. unc53

D P T D I V S K K W P V F D G E N P E N V L K R L Q L Q D L V P S

Tuesday, 18 November 1997 10:35 Page 12
fig 30 pEGFP72 (1 > 9697) Site and Sequence

```
BspM I  Xho I  Sph I
CCTGCCAACTCATCCCGACAACACTTCAATCCCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACT 6100
GGACGGTTGAGTAGGGCTGTTGTGAAGTTAGGGGAGCTCAGCAACTAGGTTAACGTACGATGGTTCGTAGTCTGGTAGCTGTTGTAAACTTGTCTTCTGA
eGFP.C.e.unc53
C.e. unc53
P  A  N  S  S  R  Q  H  F  N  P  L  E  S  L  I  Q  L  H  A  T  K  H  Q  T  I  D  N  I  .  T  E  D

Asp 718  Kpn I
CTAATCTTCTCTCGCCTCTCCCCCGCTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCCCCCTTTTCCCCCCAATTTCCCAGAACCTCCT 6200
GATTAGAAGAGAGCGGAGAGGGGGCGAAAGGAATAGAAGCATGGCCATGGACTACTAAGGGGTAAAAGGGGGAAAAGGGGGGTTAAAGGGTCTTGGAGGA
S  N  L  L  S  P  L  P  R  F  P  Y  L  R  T  G  T  .  .  F  P  I  F  P  L  F  P  P  I  S  Q  N  L  L

Xma I  Sma I  Dra I  Xmn I
GTTCCCTTTGTTCCTAGTCCTCCCGGGTGCCGACGCCGAAGCGATTTAAAAACCTTTTTCTTTCCGAAACATTTCCCATTGCTCATTAATAGTCAAATTG 6300
CAAGGGAAACAAGGATCAGGAGGGCCCACGGCTGCGGCTTCGCTAAATTTTTGGAAAAAGAAAGGCTTTGTAAAGGGTAACGAGTAATTATCAGTTTAAC
F  P  L  F  L  V  L  P  G  A  D  A  E  A  I  .  .  K  P  F  S  F  R  N  I  S  H  C  S  L  I  V  K  L

Xma I  Apa I  Sma I  Xho I  Bam HI  Xba I  Bcl I
AATAAACAGTGTATGTACTTAAAAAAAAAAAAAAAAAAAACTCGAGGGGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACA 6400
TTATTTGTCACATACATGAATTTTTTTTTTTTTTTTTTTTGAGCTCCCCCCGGGCCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGGTGT
N  K  Q  C  M  Y  L  K  K  K  K  K  K  L  E  G  G  P  G  I  H  R  I  .  I  T  D  H  N  Q  P  Y  H

Dra I  Bsm I  Hpa I
TTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAG 6500
AAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTC
I  C  R  G  F  T  C  F  K  K  P  P  T  P  P  P  E  P  E  T  .  N  E  C  N  C  C  C  .  L  V  Y  C  S

Bsm I
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGT 6600
GAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACA
L  .  V  L  Q  I  K  Q  .  H  H  K  F  H  K  .  S  I  F  F  T  A  F  .  L  W  F  V  Q  T  H  Q  C

Mlu I  Ssp I
ATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAA 6700
TAGAATTGCGCATTTAACATTCGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTT
I  L  T  R  K  L  .  A  L  I  F  C  .  N  S  R  .  I  F  V  K  S  A  H  F  L  T  N  R  P  K  S  A
```

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 13

BsrI

AATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAA
TTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCTCAGGTGATAATTTCTTGCACCTGAGGTTGCAGTTT 6800
K S L I N Q K N R P R . G . V L F Q F G T R V H Y . R T W T P T S

Dra III

GGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACC
CCCGCTTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCCAGCTCCACGGCATTTCGTGATTTAGCCTTGG 6900
G E K P S I R A M A H Y V N H H P N Q V F W G R G A V K H . I G T

Nae I

CTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGC
GATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCG 7000
L K G A P D L E L D G E S R R T W R E R K G R K R K E R A L G R W

KspI KspI

AAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
TTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGGCGGCGCGAATTACGCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCT 7100
Q V . R S R C A . P P H P P R L M R R Y R A R Q V A L F G E M C A E

BspH I Ssp I Ear I Ksp632I

ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTC
TGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCCTTCTCAG 7200
P L F V Y F S K Y I Q I C I R S . D N N P D K C F N N I E K G R V

OxaN I Pvu II Sph I Ava III Nsi I

CTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAAT
GACTCCGCCTTTCTTGGTCGACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTA 7300
L R R K E P A V E C V S V R V W K V P R L P S R Q K Y A K H A S Q

Sph I Ava III Nsi I

TAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAA
ATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATT 7400
L V S N Q V W K V P R L P S R Q K Y A K H A S Q L V S N H S P A P N

BsrI Nco I Bgl I Sfi I

CTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
GAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCG 7500
S A H P A P N S A Q F R P F S A P W L T N F F Y L C R G R G R L G

Tuesday, 18 November 1997 10:35 Page 14
fig 30 pEGFP72 (1 > 9697) Site and Sequence

```
Stu I
Avr II
Cla I
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGA 7600
GAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTCTAGCTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACT
L . A I P E V V R R L F W R P R L L Q R S I K R Q D E D R F A . L

BspM I
Xma III
ACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTG 7700
TGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCAC
N K M D C T Q V L R P L G W R G Y S A M T G H N R Q S A A L M P P C

Nar I
Bbe I
KspI
TTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGC 7800
AAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCG
S G C Q R R G A R F F L S R P T C P V P . M N C K T R Q R G Y R G

Bal I
Fsp I
Pvu II
Tth I
TGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCT 7900
ACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGA
W P R R A F L A Q L C S T L S L K R E G T G C Y W A K C R G R I S

BspM I
GTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAA 8000
CAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTT
C H L T L L L P R K Y P S W L M Q C G G C I R L I R L P A H S T T K

Ear I
Ksp632I
GCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAC 8100
CGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTG
R N I A S S E H V L G W K P V L S I R M I W T K S I R G S R Q P N

Sph I
Nco I
TGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTT 8200
ACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAA
C S P G S R R A C P T A R I S S . P M A M P A C R I S W W K M A A

Nae I
Rsr II
Ear I
Ksp632I
TTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGG 8300
AAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACC
F L D S S T V A G W V V R T A I R T . R W L P V I L L K S L A A N G
```

Tuesday, 18 November 1997 10:35

fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 1✓

GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGG

CGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCC 8400

L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S G

Asu II BspM I

GTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGG

CAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCC 8500

V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F P

Nae I Kspl Avr II

GACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGG

CTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCCTCTGTTATGGCC 8600

G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T G

Kspl

AAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTC

TTCCTTGGGCGCGATACTGCCGTTATTTTTCTGTCTTATTTTGCGTGCCACAACCCAGCAAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAG 8700

R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A L

TGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCA

ACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAAGAAGGAAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGT 8800

C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A R

Alwn I OxaN I Dra I Dra I

GCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGA

CGGTTGCAGCCCCGCCGTCCGGGACGGTATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACT 8900

S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G E

BspH I

AGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA

TCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCT 9000

D P F . . S H D Q N P L T . V F V P L S V R P R R K D Q R I F L R

TCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCG

AGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGC 9100

S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F S

Bsrl

AAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT

TTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTA 9200

E G N V L Q Q S A D T K Y C P S S V A V V R P P L Q E L C S T A Y I

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 **:**
C12N 15/12, 5/10, 15/85, C07K 14/435, 16/18, A61K 38/17, 49/00, C12Q 1/02, G01N 33/53

A3

(11) International Publication Number: **WO 98/24810**

(43) International Publication Date: 11 June 1998 (11.06.98)

(21) International Application Number: PCT/EP97/06956

(22) International Filing Date: 3 December 1997 (03.12.97)

(30) Priority Data:
9625283.8 4 December 1996 (04.12.96) GB

(71) Applicant *(for all designated States except US)*: JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B-2340 Beerse (BE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PLATTEEUW, Christ, Jules [BE/BE]; Evergemsesteenweg 17, B-9032 Wondelgem (BE). BUESA ARJOL, Carlos, Manuel [ES/ES]; Travessera de les Corts, 171/702a, E-08028 Barcelona (ES). DERAEYMAEKER, Marc [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VERHASSELT, Peter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). PUJOL, Nathalie, Jeanne, Raymonde [FR/BE]; 213, avenue du Père Soulas, F-34000 Montpellier (FR). MAERTENS, Luc, Jacques, Simon [BE/BE]; Vier Uitersten 26, B-8200 Brugge (BE). LUYTEN, Walter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). GEERTS, Hugo [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VANDEKERCKHOVE, Joel, Stefaan [BE/BE]; Rode Boukendreef 27, B-8210 Loppem (BE). GEYSEN, Johan [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). BOGAERT, Thierry, André, Olivier, Eddy [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE).

(74) Agent: BALDOCK, Sharon, Claire; Boult Wade Tennant, 27 Furnival Street, London EC4A 1PQ (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 21 January 1999 (21.01.99)

(54) Title: VERTEBRATE HOMOLOGUES OF UNC-53 PROTEIN OF C. ELEGANS

(57) Abstract

Vertebrate protein homologues of UNC-53 protein of C. elegans and nucleic acid sequences coding for said homologues or functional equivalents thereof are identified. The nucleic acid sequences in an appropriate vector are used to transfect or transform cells, tissues or organisms useful in identifying inhibitors or enhancers of the vertebrate homologue, or further proteins involved in the signal transduction pathway of which said vertebrate homologue is a component. Any of said inhibitors or enhancers identified can be included in a pharmaceutical composition or in the preparation of a medicament for treating conditions such as neurological diseases, acute traumatic injuries and to promote neuronal regeneration and inhibit metastasis or loss of contact inhibition.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 97/06956

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N5/10 C12N15/85 C07K14/435 C07K16/18
A61K38/17 A61K49/00 C12Q1/02 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 96 38555 A (BOGAERT THIERRY (BE);STRINGHAM EVE (CA); VANDEKERCKHOVE JOEL (BE)) 5 December 1996	1,2, 24-26, 28, 30-36, 38,40, 42-51, 66,68, 70,72, 78,79, 83-87, 98-105, 107,112
P,Y	see page 2, line 18 - page 20, line 26	3-23,27, 29,37, 39,41, 58,59, 67,69, 74, 80-82,
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 October 1998	27. 11. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2, NL - 2280 HV Rijswijk, Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Donath, C

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
	see figure 5 see Sequence Listing: SEQ ID NO.: 1 and 4	88, 115-118, 124,125
Y	HEKIMI, S. AND KERSHAW, D.: "Axonal guidance defects in a Caenorhabditis elegans mutant reveal cell-extrinsic determinants of neuronal morphology" THE JOURNAL OF NEUROSCIENCE, vol. 13, no. 10, October 1993, pages 4254-4271, XP000612286 see page 4254, left-hand column, paragraph 1 - page 4255, right-hand column, paragraph 2 see page 4267, right-hand column, paragraph 1 - page 4271, left-hand column, paragraph 2	3-23,27, 29,37, 39,41, 58,59, 67,69, 74, 80-82, 88, 115-118, 124,125
A	STERN, M.J. ET AL.: "The human GRB2 and Drosophila Drk genes can functionally replace the Caenorhabditis elegans cell signaling gene sem-5" MOLECULAR BIOLOGY OF THE CELL, vol. 4, no. 11, November 1993, pages 1175-1188, XP002079466 cited in the application see the whole document	

1

International application No.
PCT/EP 97/06956

INTERNATIONAL SEARCH REPORT

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 49-51, 58, 59, 107, 115 and 118 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.: 52-57,60-65,71,73,75-77,89-94,108-111,113,114
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 52-57,60-65,71,73,75-77,89-94,108-111,113,114

Claims 52-57, 60-65, 75-77, 91-94, 108-111,113 and 114 concern a compound, pharmaceutical composition, a nucleic acid sequence or the use of a compound. These compounds or compositions, however, are only defined by the method which can be used in order to identify these compounds or compositions as enhancers or inhibitors of regulation of cell shape, cell growth or motility or of the direction of cell migration or of the signal transduction pathway. Since it is completely unclear which kind of substances will be identified by the respective method and since in the specification no concrete examples for these kind of substances are given, the scope of said claims is totally ambiguous and undefined. Moreover, it cannot be excluded that even substances known in the art may be recognized as an enhancing or inhibitoring compound by the respective used method.
The same applies to claims 71, 73, 89 and 90 concerning methods comprising proteins which are only defined by a reference to another method which was used to identify said proteins.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 97/06956

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9638555 A	05-12-1996	AU 6123496 A	18-12-1996
		EP 0832222 A	01-04-1998

Form PCT/ISA/210 (patent family annex) (July 1992)